博士学位請求論文

対話における聞き手のフィードバックの評価・種類の自動推定

指導教員　宮田　章裕　教授

日本大学　大学院総合基礎科学研究科　地球情報数理科学専攻

6122D02　大西　俊輝

論文提出日　2025年1月8日

概要

本研究は，対話参加者（話し手および聞き手）の言語・非言語行動から，聞き手によるフィードバックに対する評価を推定する機械学習モデルや，聞き手のフィードバックの種類を推定する機械学習モデルを構築する取り組みを行う．具体的には，本研究は，実対話データと聞き手によるフィードバックに対する評価値を記録した対話コーパスから，話し手および聞き手の視覚的，韻律的，言語的特徴量を抽出し，褒める行為の検出と上手さの推定，理解度の推定，相槌機能の予測を行う機械学習モデルを構築する．

日常生活や社会活動において，人間同士の対面コミュニケーションは必要不可欠なものである．その中で，話し手に対する聞き手のフィードバックは，聞き手が話し手の話を聞いている合図や，聞き手が話し手の発話内容を理解している，もしくは理解しようとしている合図であると考えられている．このことから，聞き手のフィードバックは，コミュニケーションの成立や円滑な推進において，極めて重要な要素であるといえる．しかし，人とのコミュニケーションが苦手な人には，相手の話に対して適切なフィードバックが行えないという問題がしばしば生じる．これにより，コミュニケーション中に誤解や食い違いが発生し，コミュニケーションが円滑に進まなくなる可能性がある．さらには，良好な人間関係を築くことが難しくなる可能性もある．

このような問題を解決するために，心理学，教育学，社会学など，様々な分野で聞き手のフィードバックに関する研究が行われている．従来は，特定のコミュニケーションシーンを想定したアンケート調査の結果や，実際の行動を人間が観察した結果を，研究者が手作業で集計して分析するスタイルが主流であった．しかし近年，人の行動分析に適用可能な情報技術の発展に伴い，コミュニケーション中の行動をセンシングしたデータに対して，機械学習などの手段を駆使して分析を行うアプローチが多く用いられるようになってきた．これらの研究では，コミュニケーション中の人の言語・非言語行動から，特定タスクに対する評価や能力を機械学習で推定することが行われる．

上記をふまえ，本研究は，話し手および聞き手の言語・非言語行動から，聞き手によるフィードバックに対する評価を推定する機械学習モデルや，聞き手のフィードバックの種類を推定する機械学習モデルを構築する取り組みを行う．特に，円滑なコミュニケーションや良好な人間関係構築のために重要な (1) 褒める行為，(2) 理解を表す行為，(3) 相槌を打つ行為に着目し，それぞれに対応する次の 3 つの課題に取り組む．

第 1 の課題では，話し手および聞き手の言語・非言語行動から，褒める行為の検出や褒め方の上手さを推定する取り組みを行う．褒める行為は，対話相手の行動や性格に向けられた賞賛を表す言語・非言語行動であると考えられている．さらに，話し手（褒められる人）だけでなく，聞き手（褒める人）の行動やモチベーションにも影響を与えることが示唆されている．多くの既存研究では，教育やビジネスなどの特定のシーンにおける褒める行為の効果や影響を明らかにすることにとどまっており，コミュニケーション中に褒める行為を行っているか，どれくらい上手く褒めているかを自動的に検出・評価する取り組みは行われていない．そのため，聞き手は，対話相手を褒めることができているのか，どれ

くらい上手く褒めることができているのか熟練した他者の協力を得るなどしないとわからないという問題がある．この問題を解決するために，聞き手は褒める行為を行っているのか，どれくらい上手く褒めているのかを自動的に判定する必要がある．そこで第 1 の課題では，話し手および聞き手の言語・非言語行動から機械学習を用いて褒める行為の検出や褒め方の上手さの推定を行った．具体的には，実対話と褒め方に関する評価を記録した対話コーパスを構築し，視覚的，韻律的，言語的特徴量を用いて褒める行為を検出する機械学習モデル，および褒め方の上手さを推定する機械学習モデルを構築する取り組みを行った．その結果，話し手と聞き手の言語・非言語行動から褒める行為を検出できることが明らかになった．特に，聞き手の韻律的，言語的特徴量と話し手の視覚的，韻律的特徴量を用いたモデルが最も推定性能が高い結果であった．さらに，聞き手と話し手の言語・非言語行動から褒め方の上手さを推定できることも明らかになった．特に，聞き手の視覚的，韻律的，言語的特徴量と話し手の韻律的特徴量を用いたモデルが最も推定性能が高い結果であった．

第 2 の課題では，聞き手の言語・非言語行動から聞き手の理解度を推定する取り組みを行う．聞き手の理解を表す行為は，話し手の発話に対して理解を示し，コミュニケーションを促進させるために重要であると考えられている．しかし，コミュニケーションが苦手な人は，自身が理解できている，もしくは理解できていないことを対話相手に伝えられていない可能性がある．さらに，自身の発言に対して対話相手が理解できている，もしくは理解できていないことを把握できていない可能性がある．この問題を解決するためには，コミュニケーション中に聞き手が理解できているか否かを，自動的に判定する方法が有効である．そこで第 2 の課題では，聞き手の言語・非言語行動から聞き手の理解度を機械学習を用いて推定する取り組みを行った．具体的には，実対話と聞き手の理解度を記録した対話コーパスを利用し，聞き手の視覚的，韻律的，言語的特徴量から聞き手の理解度を推定する機械学習モデルを構築した．その結果，聞き手の言語・非言語行動から聞き手の理解度を推定できることが明らかになった．特に，聞き手の視覚的，韻律的，言語的特徴量を用いたモデルが最も推定性能が高い結果であった．

第 3 の課題では，話し手の言語・非言語行動から聞き手の相槌機能を予測する取り組みを行う．聞き手の相槌は，対話を成立させ，円滑に進めるために重要であると考えられている．さらに，日本語対話における相槌は，継続を促す表現，理解を示す表現，同意をする表現，感情の表現，情報を要求する表現などの機能別に分類することができると考えられている．しかし，相槌の自動評価を行う多くの既存研究では，相槌の有無やタイミングを予測する取り組みに終止しており，どのような機能を持った相槌を打つべきか予測する取り組みは行われていない．そのため，コミュニケーション中に相槌を打つ適切なタイミングは明らかになるが，打つべき相槌の機能が適切であるのか明らかでないという問題がある．この問題を解決するために，第 3 の課題では，話し手の言語・非言語行動から聞き手の相槌機能を予測する取り組みを行った．具体的には，実対話と聞き手の相槌機能を記録した対話コーパスを利用し，視覚的，韻律的，言語的特徴量から聞き手の相槌機能を予測する機械学習モデルを構築する取り組みを行った．その結果，話し手の言語・非言語行動から聞き手の相槌機能を予測できることが明らかになった．特に，話し手の視覚的，韻

律的，言語的特徴量を用いたモデルが最も予測性能が高い結果であった．

上記の 3 つの課題に取り組んだ結果から，コミュニケーション中の言語・非言語行動を機械学習を用いて分析することで，(1) 褒める行為の検出や褒め方の上手さの推定が可能であり，聞き手の韻律的・言語的特徴量と話し手の韻律的特徴量に関する行動が推定のために重要であること，(2) 聞き手の理解度の推定が可能であり，聞き手の視覚的，韻律的，言語的特徴量に関する行動が推定のために重要であること，(3) 聞き手の相槌機能を推定することが可能であり，話し手の視覚的，韻律的，言語的特徴量が推定のために重要であることが明らかになった．これらの結果は，人々が適切にフィードバックするための技術を向上させる手段や，上手にフィードバックを行う対話システムの実現に貢献しうるものである．したがって，本研究は，人々の日常生活や社会活動において，人同士の豊かなコミュニケーションの実現に貢献しうるものであると考えられる．

Abstract

This study attempts to construct machine learning models for predicting the evaluations of listener's feedback and classifying the types of listener's feedback based on the verbal and nonverbal behaviors of dialogue participants (speaker and listener). Specifically, this study extracts visual, acoustic, and linguistic features of the speaker and listener from dialogue corpuses that contains real dialogue data and the evaluations for listener ' s feedback, and constructs machine learning models that detect the praising behavior, predict the degree of praising skills, predict the degree of comprehension, and classify the feedback types.

Human communication is essential in daily life and social activities. The listener's feedback is regarded as a signal that the listener is actually listening to the speaker and the listener is either understanding or attempting to understand. Therefore, the listener's feedback is an extremely important element in establishing and facilitating smooth communication. People who lack communication skills often struggle to provide appropriate feedback during conversations, leading to misunderstandings or discrepancies that hinder the flow of communication. Additionally, building strong interpersonal relationships becomes more challenging.

To address these problems, many studies on the listener's feedback have been conducted in various research fields such as psychology, pedagogy, and sociology. Traditionally, many studies have manually tabulated and analyzed the results of questionnaire surveys that assume specific communication scenarios or the results of human observation of actual behavior. In recent years, many approaches have been used to analyze sensed data of human behavior during communication by machine learning and other methods, with the advancement of information technology applicable to the analysis of human behavior. These studies attempted to predict or classify the evaluation of specific tasks or individual abilities using machine learning. Accordingly, this study attempts to construct machine learning models for predicting the evaluation of listener's feedback and classifying the types of listener's feedback based on the verbal and nonverbal behaviors of speakers and listeners. This study focuses on the following three tasks: (1) praising behavior, (2) expressing comprehension, and (3) giving backchannels, all of which are expected to build positive human relationships and are important to facilitating smooth communication.

Regarding the first task, praising behavior is defined as verbal and/or nonverbal expression of approval that is directed toward the behavior and/or character of the target. Furthermore, this behavior has been suggested to impact the behavior and motivation of not only speaker (the person who receives praise) but also listener (the person who gives praise). Many existing studies have been limited to clarifying the effects and impacts of praising behavior in specific situations such as education and business, and no studies have been conducted to automatically detect and evaluate whether or not praising behavior is

performed during communication and how well done the praising behavior is. Because of this, there is a problem that the listener is unable to know whether he/she is able to praise the partner and how well he/she is able to do so without the cooperation of others who are experts in the field. To solve this problem, it is necessary to automatically determine whether the listener is performing the praising behavior and how well he/she is praising. Therefore, the first task aimed to detect the praising behavior and predict the degree of praising skill using machine learning based on the verbal and nonverbal behaviors of the listener and speaker. Specifically, this task focused on constructing machine learning models to detect the praising behavior and predict the degree of praising skill using the visual, acoustic, and linguistic features from a dialogue corpus containing real dialogue data and ratings of the praising behavior. The results showed that praising behavior could be detected based on the verbal and nonverbal behaviors of both the listener and speaker. In particular, the model using the acoustic and linguistic features of the listener and the visual and acoustic features of the speaker had the best prediction performance. Furthermore, the degree of praising skill could be predicted based on the verbal and nonverbal behaviors of both the listener and speaker. In particular, the model using the visual, acoustic, and linguistic features of the listener and the acoustic features of the speaker had the best prediction performance.

Regarding the second task, the listener's comprehension is important to express comprehension of the speaker's utterance and to promote communication. However, people who are not good at communication may fail to convey whether they comprehend or do not comprehend the dialogue to their partner. Furthermore, they may also be unable to grasp whether their dialogue partner understands or does not understand their own statements. To solve this problem, a method to automatically determine whether the listener comprehends during communication would be useful. Therefore, the second task aimed to predict the degree of listener's comprehension using machine learning based on the verbal and nonverbal behaviors of the listener. Specifically, this task focused on constructing machine learning models to predict the degree of listener's comprehension based on the visual, acoustic, and linguistic features of the listener using a dialogue corpus containing real dialogue data and the listener's comprehension levels. The results showed that the degree of listener's comprehension could be predicted from the verbal and nonverbal behaviors of the listener. In particular, the model using the visual, acoustic, and linguistic features of the listener had the best performance.

Regarding the third task, the listener's backchannels are important for establishing and promoting smooth communication. Furthermore, the backchannels in Japanese dialogue are classified according to their functions, such as expressions that encourage continuation, indicate understanding, agreement, emotion, and request information. However, many studies on the automatic evaluation of backchannels have focused solely on predicting the occurrence or timing of backchannels, without attempting to predict what kind

of backchannel with specific functions should be used. Therefore, although the appropriate timing for using backchannels during communication has been clarified, the issue of whether the function of the backchannels used is appropriate still remains. To solve this problem, the third task aimed to classify the types of listener's backchannels based on the verbal and nonverbal behaviors of the speaker. Specifically, this task focused on constructing machine learning models that classify the types of listener's backchannel based on the visual, acoustic, and linguistic features of the speaker using a dialogue corpus containing real dialogue data and the types of listener's backchannel. The results showed that the types of listener's backchannel could be classified based on the verbal and nonverbal behaviors of the speakers. In particular, the model using the visual, acoustic, and linguistic features of the speaker had the best performance.

From the results of the above three tasks, this study clarified the following three findings. (1) The praising behavior could be detected and the degree of praising skill could be predicted using machine learning from verbal and nonverbal behaviors, and the behaviors related to the acoustic and linguistic features of the listener and the acoustic features of the speaker were useful. (2) The degree of listener's comprehension could be predicted using machine learning from verbal and nonverbal behaviors, and the behaviors related to the visual, acoustic, and linguistic features of the listener were useful. (3) The types of listener's backchannels could be classified using machine learning from verbal and nonverbal behaviors, and the behaviors related to the visual, acoustic, and linguistic features of the speaker were useful. These results suggest that this study contributes to the improvement of techniques for people to give appropriate feedback, and to the construction of dialogue systems that give good feedback when communicating with humans. Therefore, this study is expected to contribute to fostering richer communication in people's daily life and social activities.

目 次

図 目 次

表 目 次

第1章　序論

1.1 研究の背景と目的

日常生活や社会活動において，人間同士のコミュニケーションは必要不可欠なものである．古来より，人々はコミュニケーションを通じて情報伝達や意思疎通を行い，他者との関係を築き上げてきた [4]．コミュニケーションは，言語的もしくは非言語的手段を用いて情報伝達を行うことと定義されている [5]．人々は，言葉，ジェスチャ，表情行動，視線行動，音声行動などを用いて，他者とコミュニケーションを行っている．これらの行動を適切に用いることで，人々のコミュニケーションが成立し，円滑に進むことが期待される．

コミュニケーションは，信号（メッセージ）の送信者（話し手）と受信者（聞き手）が存在し，送信者の送る信号を受信者が受け取ることで成り立つ．このとき，聞き手は話し手のメッセージを解釈すること，そして話し手は聞き手が自身のメッセージを解釈したことを把握することが重要である．コミュニケーションは，話し手と聞き手の相互関係を必要とするものと言われている [6]．その中で，聞き手が話し手に対して行うフィードバックは，コミュニケーションを成立させる上で重要な要素のひとつである [7]．聞き手のフィードバックは，聞き手が話し手の話を聞いている合図や，話し手の発話内容を理解している，もしくは理解しようとしている合図と言われている [8,9]．さらに，聞き手のフィードバックは，肯定的，矯正的，否定的の評価側面があること [10]，肯定的なフィードバックが行われると話し手の発言が増加することが報告されている [11]．これらのことから，聞き手のフィードバックが，コミュニケーションを進めるための手助けとなる [12]．ここで，聞き手がフィードバックを行う際，発話内容などの言語行動と，音声，視線，表情，ジェスチャなどの非言語行動が用いられていることがわかっている [13,14]．コミュニケーション中の聞き手の非言語行動は，聞き手の心理状態や感情的反応を表すと言われている [15]．聞き手がフィードバックを効果的に行い，コミュニケーションを円滑に進めるためには，フィードバックに関する適切な知識や技能を持ち，言語・非言語行動を適切に用いる必要があると考えられる．しかし，人とのコミュニケーションが苦手な人やコミュニケーションスキルが乏しい人は，言語・非言語行動を適切に用いてフィードバックを行うことができないという問題がある．そのため，コミュニケーション中に誤解や食い違いが発生し，コミュニケーションが円滑に進まなくなる可能性がある．さらには，良好な人間関係が築けなくなる可能性もある．

このような問題を解決するために，心理学，教育学，社会学など，様々な学術分野でコミュニケーション中の聞き手のフィードバックに関する研究が多く行われている [7,9,12,13,15–21]．これらの研究では，コミュニケーション中に聞き手がフィードバックを行うタイミング [18,19] や，フィードバックが話者交替にどのように影響するのか [17] を明らかにしている．加えて，これらの調査結果は，コミュニケーションの構造を分析し，聞き手のフィードバックがどのように行われるのか，どのような影響を与えるのかを知識として提供している．しかし，人とのコミュニケーションが苦手な人やコミュニケーションスキルが乏しい人が，自身のフィードバックの技能を把握するための取り組みや，その技能を向上させるための取り組みは行われていない．そのため，人々が自身のフィードバックの技能を把握し，さらに技能を向上させるためには，心理学者や専門家からの指導を仰ぎ，

訓練を行う必要がある．

一方，マルチモーダルインタラクション [22] や社会的信号処理 [23,24] と呼ばれる学術分野において，コミュニケーション中の行動をセンシングしたデータを用いて，機械学習などを行う研究が多く存在する．これらの研究では，心理学，教育学，社会学などの知見に基づき，人々の言語・非言語行動から特定のタスクに対する評価や個人の能力を機械学習を用いて推定・分類する取り組みが行われている [25–38]．これらは外部観測可能なデータを分析するアプローチであるため，相対的に主観に頼ることが多い心理学・教育学・社会学では得られなかった新たな知見が獲得されつつある．さらには，心理学者や専門家だけが評価・解釈できることに対して，マイク，カメラ，深度センサなどで記録した人々の行動から，機械学習を用いて評価・解釈できるようになることが期待される [23]．

上記の研究背景をふまえ，本研究は，話し手および聞き手の言語・非言語行動から聞き手のフィードバックに対する評価を推定する機械学習モデルや，聞き手のフィードバックの種類を分類する機械学習モデルを構築する取り組みを行う．この取り組みにより，聞き手のフィードバックを自動的に評価することや，適切なフィードバックを自動的に判定して提示することができ，聞き手のフィードバックに関する技能を取得するための言語・非言語行動が明らかになると考えられる．

1.2 研究の概要

1.1 節で述べたように，本研究は，話し手および聞き手の言語・非言語行動から聞き手のフィードバックに対する評価を推定する機械学習モデルや，聞き手のフィードバックの種類を分類する機械学習モデルを構築する取り組みを行う．聞き手のフィードバックは，肯定的，矯正的，否定的の評価側面に基づいて分類される [10]．肯定的な評価側面に基づくフィードバックは，話し手の発言を増加させることが報告されている [11]．このことから，円滑なコミュニケーションを実現するためには，聞き手による肯定的なフィードバックが重要であると考えられる．肯定的な評価側面を持つフィードバックとして，相槌 [9,39]，褒め（称賛）[40,41]，理解 [8,9,16]，共感 [42–44]，感謝 [45] などが挙げられる．肯定的な評価側面を持つフィードバックの中で，心理学や教育学の分野では褒める行為や相槌を打つ行為が注目されている．褒める行為は，対象の行動や性格に向けられた称賛を表現する言語・非言語行動である [46,47]．相槌を打つ行為は，話し手の発話に対して聞き手が送る短い表現であり，受容，驚き・関心，同意，正の感情表出などの機能を持つ言語・非言語行動である [9,39,48–50]．聞き手が褒める行為や相槌を打つ行為を行い，これらの行為による効果を発揮するためには，話し手の発話を理解すること [8,9,16] も重要であると考えられる．上記をふまえ，より良好な人間関係の構築が期待でき，より円滑なコミュニケーションを実現する上で重要な役割を果たす (1) 褒める行為 [40,41]，(2) 理解を表す行為 [8,9,16]，(3) 相槌を打つ行為 [9,39] に着目し，次の 3 つの課題に取り組む．

第 1 の課題では，話し手および聞き手の言語・非言語行動から褒める行為の検出や褒め方の上手さを推定する取り組みを行う [51–53]．褒める行為は，対話相手の行動や性格に向けられた賞賛を表す言語・非言語行動であると考えられている．さらに，相手を褒める

ことは，称賛の受け手（話し手）だけでなく，褒め手（聞き手）の行動やモチベーションにも影響を与えることが示唆されている．しかしながら，多くの既存研究では，教育やビジネスなどの特定のシーンにおける褒める行為の効果や影響を明らかにすることにとどまっており，コミュニケーション中に褒める行為を行っているか否かや，相手を上手く褒めている程度を，自動的に検出・評価する取り組みは行われていない．そのため，褒め手は，対話相手を褒めることができているのかや，どれくらい上手く褒めることができているのかは，専門家や経験者などの熟練した他者の協力を得るなどしないとわからないという問題がある．この問題は，褒める行為の検出・評価を自動化することで解決できる可能性がある．自動的に検出・評価ができるようになることで，熟練した他者の協力を得なくても自身の褒め方の傾向や改善するための手がかりが得られるようになると考えられる．そこで第 1 の課題では，褒め手および受け手の言語・非言語行動から機械学習を用いて褒める行為の検出や褒め方の上手さの推定を行う．具体的には，実対話と褒め方に関する評価を記録した対話コーパスを構築し，視覚的，韻律的，言語的特徴量を用いて褒める行為を検出する機械学習モデル，および褒め方の上手さを推定する機械学習モデルを構築する取り組みを行う．

第 2 の課題では，聞き手の言語・非言語行動から聞き手の理解度を推定する取り組みを行う [54]．聞き手の理解を表す行為は，話し手の発話に対して理解を示し，コミュニケーションを促進させるために重要であると考えられている．しかし，コミュニケーションが苦手な人が聞き手となる場合，自身が理解できている，もしくは理解できていないことを対話相手に伝えられていない可能性がある．さらに，そのような人が話し手となる場合，自身の発言に対して対話相手が理解できている，もしくは理解できていないことを把握できていない可能性がある．この問題を解決するためには，コミュニケーション中に聞き手が理解できているか否かを，自動的に判定する方法が有効である．そこで第 2 の課題では，聞き手の言語・非言語行動から聞き手の理解度を機械学習を用いて推定する取り組みを行う．具体的には，実対話と聞き手の理解度を記録した対話コーパスを利用し，聞き手の視覚的，韻律的，言語的特徴量から聞き手の理解度を推定する機械学習モデルを構築する取り組みを行う．

第 3 の課題では，話し手の言語・非言語行動から聞き手の相槌機能を予測する取り組みを行う [55]．聞き手の相槌は，対話を成立させ，円滑に進めるために重要であると考えられている．さらに，日本語対話における相槌は，継続を促す表現，理解を示す表現，同意をする表現，感情の表現，情報を要求する表現などの機能別に分類することができると考えられている．しかし，相槌を推定する多くの既存研究では，相槌の有無やタイミングを推定する取り組みが多く，相槌の機能を推定する取り組みは限定的である．そのため，コミュニケーション中に相槌を打つ適切なタイミングは明らかになるが，打つべき相槌の機能が適切であるのか明らかにならないという問題がある．この問題を解決するために，第 3 の課題では，話し手の言語・非言語行動から聞き手の相槌機能を予測する取り組みを行う．具体的には，実対話と聞き手の相槌機能を記録した対話コーパスを利用し，視覚的，韻律的，言語的特徴量から聞き手の相槌機能を予測する機械学習モデルを構築する取り組みを行う．

1.3 本論文の構成

本論文の構成は次のとおりである．1 章では，本論文の背景や概要を述べた．2 章では，対話における聞き手のフィードバックとマルチモーダルインタラクションに関する定義や研究事例について述べる．3 章では，本論文における問題の定義と研究課題について述べる．4 章では，褒め手および受け手の言語・非言語行動から褒める行為の検出や褒め方の上手さを推定する取り組みについて述べる．5 章では，聞き手の言語・非言語行動から聞き手の理解度を推定する取り組みについて述べる．6 章では，話し手の言語・非言語行動から聞き手の相槌機能を予測する取り組みについて述べる．最後に 7 章にて，本論文の結論を述べる．

第2章　対話における聞き手のフィードバックとマルチモーダルインタラクション

本研究では，話し手および聞き手の言語・非言語行動から聞き手のフィードバックに対する評価を推定する機械学習モデルや，聞き手のフィードバックの種類を分類する機械学習モデルを構築する取り組みを行う．本研究の位置付けを図 2.1 に示す．本研究は，コミュニケーション中の聞き手によるフィードバックに関する研究やマルチモーダルインタラクションに関する研究と関連しており，本章ではこれらの事例について紹介する．2.1 節では，人々が行うコミュニケーションの基本構成ついて説明する．2.2 節では，コミュニケーションにおける聞き手とフィードバックの定義について説明する．2.3 節では，マルチモーダルインタラクションや社会的信号処理の分野での研究事例について説明する．

<table>
<tr><td rowspan="3"></td><td colspan="8">コミュニケーション中における聞き手のフィードバック</td></tr>
<tr><td colspan="6">肯定的</td><td colspan="2">矯正的・否定的</td></tr>
<tr><td colspan="2">相槌</td><td rowspan="2">褒め
(称賛)</td><td rowspan="2">理解</td><td rowspan="2">共感</td><td rowspan="2">感謝</td><td colspan="2">相槌</td></tr>
<tr><td>機会・
タイミング</td><td>機能</td><td>機会・
タイミング</td><td>機能</td></tr>
<tr><td>心理学・教育学</td><td colspan="2">Yngve(1970)
Maynard(1986)</td><td>Brophy(1981)
Henderlongら(2002)</td><td>Schegloff(1982)
Yngve(1970)
Krautら(1982)</td><td>Mehrabianら(1972)
Deceyら(2012)
Zakiら(2012)</td><td>McCullough(2001)</td><td colspan="2">Yngve(1970)
Maynard(1986)</td></tr>
<tr><td rowspan="2">マルチモーダル
インタラクション</td><td rowspan="2">Fujieら(2005)
Haraら(2018)
Ishiiら(2021)
Morencyら(2008)
Muellerら(2015)
Truongら(2010)
Wardら(2000)
Ruede(2019)</td><td>飯塚ら(2023)
Boudinら(2024)</td><td rowspan="2">本研究</td><td>Buckinghamら(2012)
Kawaharaら(2013)
Turkら(2024)</td><td rowspan="2">Ishiiら(2018)
Tanら(2019)</td><td rowspan="2">–</td><td rowspan="2">Fujieら(2005)
Haraら(2018)
Ishiiら(2021)
Morencyら(2008)
Muellerら(2015)
Truongら(2010)
Wardら(2000)
Ruede(2019)</td><td>飯塚ら(2023)
Boudinら(2024)</td></tr>
<tr><td>本研究</td><td>本研究</td><td>本研究</td></tr>
</table>

図 2.1: 本研究の位置付け

2.1　コミュニケーション

人々はコミュニケーションで情報伝達や意思疎通を行い，他者との関係を築き上げてきた [4]．人々は，日常生活の中で 75%の時間を，他者に自分の話をしたり，他者の話を聞いたりするコミュニケーション活動に費やすと言われている [56]．すなわち，人々が生活していく上で，コミュニケーションは必要不可欠なものである．2.1.1 項ではコミュニケーションの基本原理と定義について説明し，2.1.2 項ではコミュニケーションの基本構成について説明する．

2.1.1　コミュニケーションの基本原理と定義

コミュニケーションは，幅広い領域で研究が行われており，定義も多岐にわたっている．まず，石井 [57] は，コミュニケーションの基本原理として次の 6 点を挙げている．

(1) コミュニケーションは送り手，受け手，メッセージ，チャネルなどの構成要素から成る相互行為の過程である．

(2) コミュニケーションは意識レベルと無意識レベルの両方で成立する．この原理は，表情や身振りなど無意識的にとられる非言語的行動があることからも明らかである．

(3) コミュニケーションは不可逆的である．この原理によると，一旦発信されて受け手に受信されたメッセージを元に戻すことは不可能である．いくら訂正したところで，それは新しいメッセージでしかない．

(4) コミュニケーションは動的である．この原理は，コミュニケーションが静止状態ではなく，常に変化しつつある状態であることを示す．

(5) コミュニケーションは組織的である．この原理は，コミュニケーションに様々な要素や条件が関わっているが，それらは有機的に関連し合って全体として組織的に作用することを示す．

(6) コミュニケーションは適応の性格をもつ．この原理によると，コミュニケーションに関わる人間は，構成要素（特に相手）や条件（例えば，その場の状況）などに適応しようと努力する．

次に，アメリカ心理学会（American Psychological Association）が，100 名以上の研究者や医師らからなる編集委員の指導のもと作成した心理学辞典 [5] によると，コミュニケーションは次のように定義されている．

> コミュニケーションは，言語的もしくは非言語的手段による情報の伝達のこと．

さらに，McCroskey ら [4, 58] は，コミュニケーションを次のように定義している．

> コミュニケーションは，ある人が言語・非言語メッセージによって他者（や人々）の心に意味を生じさせるプロセスである．

このようにコミュニケーションの定義は多岐にわたるが，言語・非言語行動を用いて他者との情報伝達・意思疎通を行うことがコミュニケーションの基本概念であるという点で既存研究の意見は一致していると考えられる．

2.1.2　コミュニケーションの基本構成

深田 [1] は，コミュニケーションは（1）送り手,（2）メッセージ,（3）チャネル,（4）受け手,（5）効果の基本構成（図 2.2）から成り立つと述べている．送り手（話し手）の役割は，伝達したい情報を言語・非言語符号に変換（符号化）することである．メッセージは，送り手によって符号化された言語・非言語符号の集合体である．送り手が伝達したい情報は，メッセージとして符号化されなければ受け手に知覚されることはないと言われている．メッセージには，言語的行動である言葉はもちろんのこと,（1）外見的特徴,（2）ジェスチャと動作,（3）表情と視線行動,（4）音声行動,（5）空間,（6）接触,（7）沈黙,（8）時間,（9）色彩,（10）匂いなどの非言語的行動も含まれている [59, 60]．チャネルは，メッセージが運ばれる経路であり，受け手がメッセージを知覚するために使用する感覚器官（視覚，聴覚，触覚，嗅覚，味覚）に基づいて，視覚的チャネルや聴覚的チャネルなど

と呼ばれている．受け手（聞き手）の役割は，送り手が符号化したメッセージをある特定のチャネルを通して受け取り，メッセージの意味を解釈することである．効果は，メッセージによって受け手が送り手に及ぼす影響のことである．人々は，上記のコミュニケーションの基本構成に則り，言語的・非言語的行動を駆使して，他者とコミュニケーションを行っている．

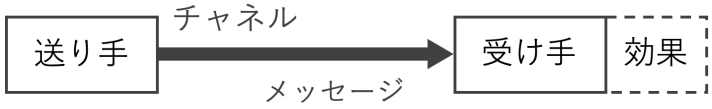

図 2.2: 深田によるコミュニケーションの基本構成 [1]

2.2　コミュニケーションにおける聞き手とフィードバック

2.1.2 項で述べたように，コミュニケーションは，メッセージの送り手（話し手）と受け手（聞き手）が存在し，送り手のメッセージを受け手が受け取ることで成り立つ．さらに，コミュニケーションは，話し手と聞き手の協同作業であると言われており [61]，話し手のメッセージを聞き手が解釈し，聞き手が解釈したことを話し手が把握することが重要であり，話し手と聞き手の相互関係を必要とするものである [6]．すなわち，話し手と聞き手が互いに反応し，コミュニケーションの方向性を調整しながら発展させることが期待されているのである．

続く各項では，聞き手の主な役割であるフィードバックについて説明する．2.2.1 項では，聞き手のフィードバックの定義について説明する．2.2.2 項では，聞き手のフィードバックの役割について説明する．2.2.3 項では，聞き手のフィードバックの種類について説明する．2.2.4 項では，聞き手のフィードバックに関する研究事例について説明する．

2.2.1　聞き手のフィードバックの定義

コミュニケーションにおける聞き手は，ただ話し手の話を聞いているだけではなく，話し手のメッセージに反応し，フィードバックを与える必要がある．フィードバックは，送り手（話し手）が発信したメッセージに対する受け手（聞き手）の評価や受容度といったメッセージの効果性を直接反映する情報の返還のことであり，メッセージの一種である（図 2.3）[2]．Bavelas ら [13] は，聞き手のフィードバックを次のように定義している．

> *Actions [both verbal and nonverbal] that indicate the person is attending, following, appreciating, or reacting to the story. They include (but are not limited to) nodding, “mhm,” “yeah,” smiling, laughing, motor mimicry, gesturing the*

content of the story, supplying words or phrases, dramatic intakes of breath, and displays of excitement, fear, or alarm.

このことから，聞き手のフィードバックは，言語・非言語行動を用いて，コミュニケーションに参加していること，話し手の話に反応していることを表すものであると考えられる．

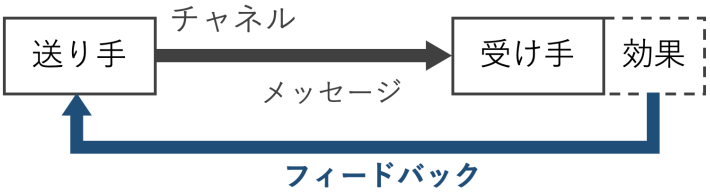

図 2.3: DeVito によるコミュニケーションのプロセスモデル（一部）[2]

2.2.2　聞き手のフィードバックの役割

聞き手のフィードバックは，バックチャネル，聞き手応答，聞き手反応とも呼ばれ，心理学，教育学，社会学など，様々な分野で研究が行われている [7,9,12,13,15–21]．聞き手のフィードバックは，聞き手が話し手の話を聞いていることを表す合図や，話し手の発話内容を理解している，もしくは理解しようとしていることを表す合図であり [8,9]，コミュニケーションを進めるための手助けとなると言われている [12]．Yngve [9] は，コミュニケーション中に，聞き手が話し手のメッセージに反応して行う短いフィードバックをバックチャネルと定義した．Duncan と Fiske [62] は，バックチャネルは聞き手がコミュニケーションに積極的であることを表し，話し手と聞き手の行動調整を促進するものであると示した．このことから，コミュニケーションを成立させるためには，聞き手のフィードバックは重要な役割を持つと考えられる．

2.2.3　聞き手のフィードバックの種類

コミュニケーション中における聞き手のフィードバックは，(1) 接触（相手が対話を続ける意思と能力があるかどうか），(2) 知覚（相手がメッセージを知覚する意思と能力があるかどうか），(3) 理解（相手がメッセージを理解する意思と能力があるかどうか），(4) 態度的反応（相手がメッセージに反応し，適切に応答する意思と能力があるかどうか，特にそれを受け入れるか拒否するか）の 4 つの機能を持っていると言われている [12]．さらに，聞き手のフィードバックは肯定的なフィードバックと否定的なフィードバックに区別できると言われている [6,12]．Verplanck [11] は，聞き手が話し手の発言に対して賛成や要約して言い換えるなどの肯定的なフィードバックを積極的に行うようになると，話し手の発言が増加することを示した．一方で，聞き手が話し手の発言に対して肯定的なフィー

ドバックを積極的に行わないようになったり，否定的な反応を行うようになったりすると，話し手の発言が減少することも示した．これらのことから，聞き手が積極的に肯定的なフィードバックを行うことで，コミュニケーションが円滑に進むことが期待される．

コミュニケーション中における肯定的なフィードバックには，相槌 [9, 39]，褒め（称賛）[40, 41]，理解 [8, 9, 16]，共感 [42–44]，感謝 [45] などが挙げられる．相槌は，話し手の発話に対して聞き手が送る短い表現であり，受容，驚き・関心，同意，正の感情表出などの機能を持つ行為である [9, 39, 48–50]．褒め（称賛）は，対象の行動や性格に向けられた称賛を表現する行為である [46, 47]．理解を表す行為は，話し手の発話に対して理解している，もしくは理解しようとしていることを表す行為である [8, 9]．共感を表すは，他者との感情共有や他者への同情喚起のような感情的側面，他者の信念の推論のような認知的側面を持つ行為である [43] [44]．感謝は，他者の道徳的な行為に対して反応する行為である [45].

2.2.4　聞き手のフィードバックに関する研究事例

教育学，心理学の分野における褒める行為に関連する研究事例の多くは，アンケート調査や特定のタスク後に実施した評価の結果を用いて分析を行い，褒める行為の効果や影響を明らかにしている [41, 63–67]．Anderson ら [68] は，絵画を描くという課題を用いて実験を行い，ポジティブな言語的報酬（例：その絵は本当にいいね）を与えられた場合は絵を描く時間が長くなり，受賞的な報酬や金銭的な報酬の場合を与えられた場合は絵を描く時間が短くなることを明らかにしている．しかし，褒める行為による効果は確認できるが，言語・非言語行動をどのように用いて褒めると効果があるのかは明らかにされていない．Stipek ら [69] は，パズルを成功させる課題に取り組んでもらったあと，「本当に上手にパズルができたね」のように成功したことについて褒める，もしくは「パズルが終わったね」のように成功した事実についてコメントするという実験を行った．実験の結果，成功した事実についてコメントされた参加者よりも，成功したことについて褒められた参加者は笑顔や堂々としたポーズなどのポジティブな表出が多くなる傾向が示唆された．

2.3　マルチモーダルインタラクション

2.1.2 項で述べたように，コミュニケーションは言語・非言語行動を用いて行われている．コミュニケーション中の人間の言語・非言語行動から得られる情報を統合し，人間の情動，性格，スキルといった社会性を理解および計算する取り組みが，マルチモーダルインタラクションや社会的信号処理と呼ばれる学術分野で行われている [23, 24]．これらの分野では，心理学，教育学，社会学などの知見に基づき，人々の言語・非言語行動から人間の感情，性格，スキル等に対する判定や評価を機械学習を用いて推定・分類する試みが多い．具体的には，感情 [25]，個人特性 [26–28]，コミュニケーション [29, 30]，プレゼンテーション [31–33]，リーダシップ [34]，インタビュー [35]，共感 [36, 37]，自己開示 [38]

に関する人間の情動，性格，スキルの推定が行われている．さらに，コミュニケーション中における聞き手のフィードバックを対象とする研究も多く存在する [70–80]．これらは外部観測可能なデータを分析するアプローチであるため，相対的に主観に頼ることが多い心理学・教育学・社会学では得られなかった新たな知見が獲得されつつある．さらには，心理学者や専門家だけが評価・解釈できることに対して，マイク，カメラ，深度センサなどで記録した人々の行動から，機械学習を用いて評価・解釈できるようになることが期待される [23].

続く各項では，マルチモーダルインタラクションの分野で取り組まれている研究事例について説明する．2.3.1 項では，コミュニケーション参加者の性格特性や能力を推定・分析する研究事例について説明する．2.3.2 項では，コミュニケーション中の聞き手のフィードバックを推定・分析する研究事例について説明する．

2.3.1　コミュニケーション参加者の性格特性や能力を推定・分析する研究

2.3.1.1　個人の性格特性を推定する研究事例

本目では，人間の言語・非言語行動を利用して，個人の性格特性を推定する研究事例について述べる [26] [27] [28]．これらの研究事例では，人間の言語・非言語行動を記録したコーパスを利用もしくは作成し，個人の性格特性を推定するための機械学習モデルを構築している．

Batrinca ら [26] は，human–machine interaction（HMI）と human–human interaction（HHI）の 2 つのシナリオで収集した音声・映像データから，Big Five の性格特性（外向性，誠実性，情緒不安定性，開放性，調和性）を自動認識している．HMI と HHI という 2 つの異なるシナリオにおける性格特性の現れ方を分析するために，この研究ではマルチモーダルデータセットを構築している．このデータセットは，Anderson ら [81] によって導入されたマップタスクコーパスを改良したものである．マップタスクは，心理学，社会学，言語学の研究で広く使われている協力的なタスクであり，タスクによって引き出された自発的な対話やインタラクションを収集するために用いられる．1 つ目のシナリオ（HMI）では，実験参加者に実験者の操作するシステムと対話を行ってもらうことで，データを収集していた．一方で，2 つ目のシナリオ（HHI）では，実験参加者に他の人と直接対話を行わせることで，データを収集していた．この研究では，2 つのシナリオのデータセットから視覚，音声（アノテーションの抽出による音声と自動抽出による音声）に関する特徴量が抽出され，性格特性を推定する機械学習モデルが構築された．その結果，1 つ目のシナリオでは正解率が 0.67，2 つ目のシナリオでは正解率が 0.70 であった．Batrinca らの結果は，性格特性の自動認識の実現可能性があることを示している．

Valente ら [27] は，自発的な会話における話者の性格特性を推測するためのアノテーションと検証を行っている．この研究では，データセットとして AMI コーパス [82] が利用された．このコーパスは，会議室で撮影されたミーティングのコーパスであり，ミーティング参加者それぞれの音声と映像が記録されている．映像には，ミーティングの全体像と参

加者をクローズアップした映像の両方が記録されている．加えて，コーパスにはシナリオに沿ったミーティングと非シナリオのミーティングが含まれている．シナリオミーティングでは，4 人の参加者がプロジェクトマネージャ，マーケティングエキスパート，ユーザインタフェースデザイナー，インダストリアルデザイナーからなるチームとして新しいリモコンをデザインするシナリオが採用されている．プロジェクトマネージャには，他の発言者と議論すべきいくつかの項目を含むアジェンダに従って会議を進行する役割が与えられている．このコーパスには，役割，発言時間，単語，対話行為に関する情報が記録されている．上記のコーパスから，彼らは言語特徴量（単語の n-gram，対話行為）と非言語特徴量（韻律，発話の活性度，発話の重複，発話のポーズ）を抽出し，ブースティングを用いて性格特性を推測する取り組みを行っている．その結果，外向性，誠実性，情緒不安定性は，それぞれ 74.5%，67.6%，68.7%の精度で推測できることが明らかになった．加えて，非言語特徴量は，言語特徴量よりも性格特性の推定に有用であることも明らかになった．

Jayagopi ら [28] は，対面での相互作用における発話と視線のパターンを特徴づける集団行動の手がかりを定義し，集団行動のパターンがメンバ間のパフォーマンスや自己および他者に対する認識に影響するのか明らかにするためのフレームワークを提案している．この研究では，ELEA コーパス（Emergent LEAder Corpus）[83] における 18 のグループインタラクションをデータセットとして用いている．ELEA コーパスは，ミーティングにおける個人の性格特性やリーダシップに関する情報と 40 件（約 10 時間）のミーティングを映像と音声で記録しているコーパスである．上記のコーパスから，彼らはグループにおける非言語行動の抽出を行い，グループでの社会的・心理的構成（性格特性，対人認識，能力等）と関連づけることを行っている．非言語行動として，個人に関する特徴量（発話状態，発話長，発話ターン，割り込み成功，割り込み失敗，バックチャネル，頭部姿勢，視線方向），グループに関する特徴量（ミーティングの継続時間，発話回数，無音区間，発話の重複区間，発話の非重複区間）が抽出されている．上記の特徴量を離散化するために，彼らは Bag of nonverbal patterns を定義し，Latent Dirichlet Allocation トピックモデルの構築を行っている．その結果，グループに関する特徴量とトピックの両方が，すべての社会的・心理的構成と有意な相関を持つことが示された．

2.3.1.2　個人の能力を推定する研究事例

本目では，人間の言語・非言語行動を利用して，個人の能力を推定する研究事例について説明する．多くの研究では，個人の能力としてプレゼンテーション [31–33]，コミュニケーション [29,30]，リーダシップ [34]，面接 [35]，共感 [36,37]，自己開示 [38] に着目している．これらの研究は，人間の言語・非言語行動を記録したコーパスを利用もしくは作成し，個人の能力を推定する機械学習モデルを構築している．

Ramanarayanan ら [31] は，人間が評価したプレゼンテーション能力のスコアを予測する取り組みを行っている．この研究は，プレゼンテーション中の 3D 身体動作，音声，映像を記録した 56 個のプレゼンテーションデータから，頭部方向，視線方向，表情，運動

量や手の動きの程度を特徴量として抽出し，プレゼンテーション能力のスコアを予測する機械学習モデルを構築している．その結果，すべての特徴量を組み合わせた機械学習モデルが最も良い予測結果であることが示された．

Chen ら [32] は，プレゼンテーションを評価するためのマルチモーダルスコアリングモデルを提案している．この研究は，プレゼンテーション中の 3D 身体動作，音声，映像を記録した 56 個のプレゼンテーションデータから，発話内容（動詞句，T-units，節など 9 種類の構文構造の頻度，平均節長，T-units あたりの動詞句など 14 種類の構文複雑度），韻律的行動（発話速度，韻律変化，ポーズ，発音），視覚的行動（運動量，手の動き，頭の向き）に関する特徴量を抽出し，プレゼンテーションの評価を予測する回帰モデルを構築している．機械学習モデルは，Random Forests（RF）と Support Vector Machine（SVM）の 2 種類を用いている．その結果，どちらのモデルにおいても，発話内容，韻律的行動，視覚的行動に関する特徴量が，プレゼンテーションの評価を予測するために有用であることが示された．加えて，3 つの特徴量のうち韻律的特徴量が予測に最も有用であること，RF モデルでは発話内容と韻律的行動が組み合わさることで予測性能が向上することが示唆された．

Yagi ら [33] は，言語・非言語行動からプレゼンテーションスキル推定するためのインスタンスの重み付けの適用法を提案している．はじめに，この研究はマルチモーダルプレゼンテーションデータセットを構築している．このデータセットは，58 のプレゼンテーションセッションにおける音声，身体動作，テキストデータと人事経験のある 2 人の外部評価者によって評価されたスコアを記録している．プレゼンテーションスキルを推定するために，音声，身体動作，言語に関する特徴量を抽出している．次に，この研究はマルチモーダル・インスタンスの重み付け適応に基づいた回帰モデルに対して，入力値を結合するモデル（Early fusion）と出力値を結合するモデル（Late fusion）の 2 つのアプローチを提案している．実験の結果，インスタンス重み付け適応を用いた入力値を結合する回帰モデルは，プレゼンテーションの目標要素の明確さに対する回帰精度を示す相関係数が改善している．

Park ら [29] は，言語的行動と非言語的行動を用いて，オンライン・ソーシャル・マルチメディア・コンテンツにおける話者の説得力を予測している．この研究は，商品レビューの動画と商品の評価のデータセットを用いている．各商品レビューには，特定の商品について話す話者の動画と，その商品に対する話者の評価である星 1 つ（最も否定的な評価）〜星 5 つ（最も肯定的な評価）が記録されている．データセットは，1,000 本のレビュー動画（肯定的なレビュー動画 500 本，否定的なレビュー動画 500 本），特定の映画について説明する話者を正面から撮影した映像データが記録されている．この研究は，上記のデータセットから，視覚（感情，ポジティブ/ネガティブの値，Action Units [84]，視線移動，姿勢），韻律（ピッチ，フォルマント，声質，Mel-frequency cepstral coefficients），言語（発話率，ポーズ，ポーズフィラー，発話の妨害率，発話の吃り）に関する特徴量を抽出し，話者の説得力を予測する機械学習モデルを構築している．その結果，3 つのモダリティを組み合わせて学習した予測モデルは，ユニモーダル予測モデルと比較して統計的に有意に良い結果を示した．

Okada ら [30] は，言語・非言語行動からコミュニケーションスキルを推定する取り組みを行っている．この研究は，コミュニケーションスキルを推定するために，MATRICS コーパス [85] を用いている．MATRICS コーパスは，4 名 1 組のグループによるディスカッションが 10 組分記録されている．ディスカッションは，対象とする人物などの順位付けに関するタスク（Task 1），事前に情報が共有された企画タスク（Task 2），事前に情報が共有されていない企画タスク（Task 3）が採用されている．MATRICS コーパスには，カメラ，マイク，深度センサなどを用いて記録したマルチモーダルデータとコミュニケーションスキルに関するスコアデータが含まれている．この研究は，上記のコーパスから言語行動（品詞，対話行為），非言語行動（発話ターン，韻律情報，頭部行動）に関する特徴量を抽出し，コミュニケーションスキルを推定するために有用であるのか分析している．具体的には，この研究はコミュニケーションスキルのスコアを推定する回帰モデルを構築し，タスクごとに最も推定性能が高い特徴量の組み合わせを明らかにしている．その結果，Task 1 では対話行為に関する特徴量が有用であること，Task 2 では韻律的特徴量が有用であること，Task 3 では視覚的特徴量が有用であった．

Nguyen ら [35] は，Hirability における応募者と面接官の非言語行動を用いて，面接での採用可能性を予測するためのフレームワークを提案している．Hirability は，職種，面接の内容，面接の行われ方に依存する社会的構成の概念である [86]．この研究は，62 件（計 670 分）の採用面接をカメラとマイクで記録したデータと Hirability スコアが記録されているコーパスを用いている．Hirability スコアは，構造化された面接の 4 つの行動（コミュニケーション能力，説得力，誠実な仕事振り，ストレスへの耐性）に関する質問による 4 つのスコアと，面接の全過程に関連する 1 つのスコアから構成されている．この研究は，上記のデータセットから，韻律的（発話のポーズ，発話時間，発話の流暢さ），視覚的（頷き，頭部の動き，視線，表情，外見），韻律と視覚の組み合わせ（音声のバックチャネル，視覚的バックチャネル，音声と視覚的のバックチャネル，相互発話，同時頷き）に関する特徴量を抽出し，応募者の採用可能性を予測する機械学習モデルを構築している．その結果，この研究は応募者と面接官の韻律的特徴量を用いたモデルが最も予測性能が高いことを明らかにしている．

Soleymani ら [38] は，対話中の言語・非言語行動から自己開示レベルを推定している．この研究は，人間同士もしくは人間とエージェントの擬似モノローグ対話が行われている Distress Analysis Interview Corpus [87] と Social Anxiety and Self-disclosure Dataset [88] の 2 つのデータセットを用いている．これらのデータセットは，外部参加者がラベル付けした自己開示に関する評価が記録されている．これらのデータセットから，この研究は視覚的，韻律的，言語的特徴量を抽出し，自己開示レベルを推定する回帰モデルを構築している．その結果，コーパスごとの評価では，この研究は言語的特徴量が自己開示の推定に最も適していることを明らかにしている．一方で，コーパス間での評価では，この研究は視覚的，韻律的特徴量が安定した性能を発揮することを明らかにしている．視覚的，韻律的，言語的特徴量を合わせたマルチモーダルモデルは，コーパスごとの評価では最良のモダリティ（言語）と同等の性能となり，コーパス間での評価では最良のモダリティを上回る性能を示した．

2.3.2　コミュニケーション中の聞き手のフィードバックを推定・分析する研究

2.3.2.1　聞き手の相槌を推定・分析する研究

様々な視点から，聞き手の相槌を推定・分析する研究が行われている．多くの研究は，コミュニケーション中の聞き手の相槌の有無やタイミングを予測する取り組みを行っている [70–78,80]．韻律的特徴量は，コミュニケーション中の相槌やバックチャネルを分析するために有用であると報告されている [9]．このことから，相槌やバックチャネルの有無やタイミングを予測する研究の多くは，韻律的特徴量に着目している．

Mueller ら [74] は，ニューラルネットワークに基づいてバックチャネルの有無を予測するアプローチを提案した．彼らの結果は，より多くのレイヤーを持つネットワークを使用することで，モデルのパフォーマンスが向上することを示している．

Ward ら [77] は，日本語におけるバックチャネルのタイミングを予測する際，沈黙の発生を待つことよりも，聞き手の音声が低いピッチとなる方が予測に有用であることを示唆している．

2.2.1 項で述べたように，聞き手のフィードバックは言語・非言語行動で表現される．このことから，韻律的特徴量だけでなく，言語的特徴量 [72,78–80] や視覚的特徴量 [72,73,80] に着目して，バックチャネルの有無やタイミングを推定する取り組みも行われている．

Morency ら [73] は，隠れマルコフモデルや条件付き確率場といった順序的確率モデルを用いて，人間同士のインタラクションデータベースから学習し，話者のマルチモーダルな特徴量（韻律，発話内容，視覚）を利用して，聞き手のバックチャネルのタイミングを予測する取り組みを行っている．彼らは，視覚的バックチャネルの予測において，手作業で作成されたルールに基づく従来の手法よりも統計的に有意に改善したことを示している．

Ruede ら [78] は，ピッチや音量といった韻律的特徴量に基づいて，ニューラルネットワークでバックチャネルの有無を予測するアプローチを提案している．LSTM モデルを用いて予測した結果，F 値は 0.375 であったことが報告されている．さらに，Word2Vec [89] による単語埋め込みを追加した結果，F 値は 0.390 に向上したことが報告されている．これらの結果は言語的特徴量を追加することで，韻律的特徴量のみを使用した予測モデルに比べて性能が向上することを示している．

Ishii ら [72] は，マルチモーダル情報（視覚的，韻律的，言語的情報）を用いて，バックチャネルを予測する取り組みを行っている．さらに，話者交替や発話意欲を同時に予測することがバックチャネルの予測に影響を与えるのか明らかにする取り組みも行っている．その結果，話し手と聞き手のマルチモーダル情報からバックチャネルが予測できること，発話意欲とバックチャネルを同時に予測することでバックチャネル予測の精度が向上することが示されている．

2.3.2.2　聞き手の理解度を推定・分析する研究

医療現場，ポスター発表などの特定シーンにおける理解度を推定する取り組みが行われている [90–92].

Buckingham ら [90] は，インフォームド・コンセントのプロセスにおける参加者の表情から彼らの理解度を推定する取り組みを行っている．

Kawahara ら [91] は，ポスター発表における聴衆の相槌や視線情報から，聴衆の発話時の理解度を推定できることを明らかにしている．

Türk ら [92] は，ビデオ回顧シーンにおける聞き手の言語・非言語行動から，聞き手の理解/非理解を推定する取り組みを行っている．彼らは韻律的および視覚的特徴量が理解/非理解の推定に寄与していることを明らかにしている．

2.3.2.3　聞き手の共感表現を推定・分析する研究

コミュニケーション中の話者の共感表現を推定する取り組みが行われている [36, 37].

Ishii ら [36] は，Davis の対人反応性指標に基づく共感スキルレベルに応じて，話者交替/継続における視線行動と対話行為について分析している．はじめに，この研究は，4 名 1 組のグループによるディスカッションにおける言語・非言語行動と共感スキルに関する評価値が記録されたマルチモーダルコーパスを構築している．次に，この研究は，共感スキルが高い人と低い人では，対話行為のカテゴリ（情報提供，自己開示，共感，話者交替，その他）における発話を伴う話者交替/継続の視線行動が定量的に異なるのか分析している．その結果，共感スキルの低い参加者は，対話行為のカテゴリが自己開示の場合，話者継続時に聞き手とのアイコンタクトを避ける傾向があり，対話行為のカテゴリが共感の場合，聞き手とのアイコンタクトを開始する傾向が見られている．共感スキルの高い参加者は，話し手の発話の対話行為のカテゴリが提供や共感の場合，話者交替時に話し手から目を逸らすことが多く，話し手の発話の対話行為のカテゴリが展開の場合，共感スキルの違いによる視線の動きの違いは見られなかった．次に，この研究は，Gaze transition patterns (GTP) と対話行為の情報を用いて，共感スキルレベルを推定するモデルを構築している．GTP とは，任意の時間区間内で，現話者，非話者，人物以外の注視対象に対して，どのような順序で視線を向けたか，また人物を注視した際に視線交差が起きたかという時系列的な遷移を N-gram パターンで表現したものである．この研究は，GTP と対話行為の情報の有用性を比較するために，発話時間や発話ターン数などの発話情報を用いた推定モデルと単純な視線情報（議論の際に話し手や聞き手を見ている時間）を用いた推定モデルを構築している．さらに，この研究は，マルチモーダル特徴量の有用性を評価するために，GTP と対話行為を用いた推定モデルと GTP，対話行為，発話情報，単純な視線情報を用いた推定モデルを構築している．その結果，多人数での議論における個人の共感スキルレベルの推定は，話者交替/継続時の視線行動と対話行為の情報が有効であることが示唆された．このことから，この研究は，話者交替/継続時の詳細な視線行動を特徴量とすることで，共感スキルレベルを推定できる可能性を示している．

Tan ら [37] は，共感している聞き手が，話し手のストーリーを聞いているときの感情の価数を時間経過とともに予測する Long Short Term Memory (LSTM) モデルを提案している．この研究は，1 人が演者（話し手）で，もう 1 人が参加者（聞き手）である 2 者対話をデータセットとして用いている．2 者対話は，話し手が聞き手に対して架空の自伝的な物語（幼なじみの話や飛行機で嫌な思いをした話など）を話すタスクが行われており，各セッションの後，聞き手は話し手がストーリーを話しているときに自身がどう感じたか（共感的反応）を −1（否定的）〜1（肯定的）の範囲で評価した．この研究は，上記のデータセットから視覚的，韻律的，言語的特徴量を抽出し，聞き手の共感的反応を予測する LSTM モデル構築している．その結果，言語的特徴量を含むモデルは，言語的特徴量含まないモデルよりも予測性能が優れていることが明らかになり，相手の話に共感する際の反応に影響を与える最も重要な情報は相手の話の内容であることが示唆された．

第3章　聞き手フィードバックの評価・種類の自動推定

日常生活や社会活動において，人々はコミュニケーションを通じて情報伝達や意思疎通を行い，他者との信頼関係を築き上げてきた．コミュニケーションは，言語的もしくは非言語的手段による情報伝達のことと定義されており [5]，人々は言葉，ジェスチャ，表情行動，視線行動，音声行動等を用いて，他者とコミュニケーションを行っている．

2.1 節で述べたように，コミュニケーションは話し手のメッセージを聞き手が受け取ることで成立する．その中で，聞き手が話し手に対して行うフィードバックは，コミュニケーションを成立させる上で重要な要素のひとつである [7]．聞き手のフィードバックは，聞き手が話し手の話を聞いていることや，その内容を理解（しようと）していることを表す合図と言われている [8,9]．これらのフィードバックを，聞き手が話し手に対して行うことで，コミュニケーションが成立し，円滑に進むことが期待される．しかし，人とのコミュニケーションが苦手な人やコミュニケーションスキルが乏しい人は，相手の話に対して適切にフィードバックを行うことができない可能性がある．フィードバックが苦手な聞き手は，話し手のことを上手く褒められなかったり，話が理解できなかったことを伝えられなかったり，適切な相槌が打てなかったりするだろう．そのとき，話し手は，自身の努力が認められなかったと感じたり，相手が理解していると勘違いして話を進めたり，スムーズに話を続けられなかったりするだろう．すなわち，聞き手がフィードバックを苦手とする場合，コミュニケーションが円滑に進まず，情報伝達や人間関係構築などの目的が十分に達成できない可能性が高まってしまう可能性がある．

心理学，教育学，社会学など，様々な分野でコミュニケーション中の聞き手に着目し，聞き手のフィードバックに関する研究が多く行われている [7,9,12,13,15–21]．これらの研究の多くは，コミュニケーション中に聞き手がフィードバックを行うタイミング [18,19] やフィードバックが話者交替にどのように影響するのか [17] を明らかにしている．コミュニケーション中の聞き手の非言語行動は，聞き手の心理状態や感情的反応を表している [15]．これらの研究の多くは，コミュニケーションの構造を分析して，聞き手のフィードバックがどのように行われるのか，どのような影響を与えるのかを知識として提供している．

2.3 節で述べたように，マルチモーダルインタラクションや社会的信号処理と呼ばれる学術分野において [23,24]，心理学，教育学，社会学など，様々な分野で得られる知見に基づいて，コミュニケーションや特定シーンにおける人間の言語・非言語行動から得られる情報を統合し，人間の情動，性格，スキルといった社会性を理解および計算する取り組みが行われている．具体的には，コミュニケーションや特定シーンにおける感情 [25]，個人特性 [26–28]，コミュニケーション [29,30]，プレゼンテーション [31–33]，リーダシップ [34]，インタビュー [35]，共感 [36,37]，自己開示 [38] に関する人間の情動，性格，スキルを推定する取り組みが挙げられる．さらに，コミュニケーション中の聞き手の理解度を対象とする研究 [92] や，フィードバックを対象とする研究 [50,70–80] も行われている．

これらをふまえ，本研究では，コミュニケーション中の話し手および聞き手の言語・非言語行動から，聞き手のフィードバックに対する評価を推定する機械学習モデルや，聞き手のフィードバックの種類を推定する機械学習モデルを構築する取り組みを行う．この取り組みにより，聞き手のフィードバックを自動的に評価することや，適切なフィードバックを自動的に判定して提示することができ，聞き手のフィードバックに関する技能を取得

するために必要となる言語・非言語行動が明らかになると考えられる．聞き手のフィードバックは，肯定的，矯正的，否定的の評価側面に基づいて分類される [10]．肯定的な評価側面に基づくフィードバックは，話し手の発言を増加させることが報告されている [11]．このことから，円滑なコミュニケーションを実現するためには，聞き手による肯定的なフィードバックが重要であると考えられる．肯定的な評価側面を持つフィードバックとして，相槌 [9,39]，褒め（称賛）[40,41]，理解 [8,9,16]，共感 [42–44]，感謝 [45] などが挙げられる．肯定的な評価側面を持つフィードバックの中で心理学や教育学の分野では褒める行為や相槌を打つ行為が注目されている．聞き手が褒める行為や相槌を打つ行為を行い，これらの行為による効果を発揮するためには，話し手の発話を理解していることも重要であると考えられる．本研究は，より良好な人間関係の構築が期待でき，より円滑なコミュニケーションを実現する上で重要な役割を果たす (1) 褒める行為 [40,41]，(2) 理解を表す行為 [8,9,16]，(3) 相槌を打つ行為 [9,39] に着目する．

3.1　聞き手のフィードバックにおける既存研究の問題点

3.1.1　褒める行為に関する既存研究の問題点

褒める行為は，対象の行動や性格に向けられた称賛を表現する言語・非言語行動であると考えられている [46,47]．さらに，褒める行為は，褒める人から褒められる人への一方的な意思の伝達ではなく，複雑な社会的コミュニケーションであり，褒められる人の役割は褒める人の役割と同じくらい重要であると考えられている [41]．2.2.4 項で述べたように，教育学，心理学の分野における褒める行為に関連する研究事例の多くは，アンケート調査や特定のタスク後に実施した評価の結果を用いて分析を行い，褒める行為の効果や影響を明らかにしている [41,63–69]．ただし，多くの既存研究では，教育やビジネスなどの特定のシーンにおける褒める行為の効果や影響を明らかにすることにとどまっており，コミュニケーション中に褒める行為を行っているか否かや，どれくらい上手く褒めているかの程度を自動的に検出・評価する取り組みは行われていない．そのため，聞き手は，対話相手を適切に褒めることができているのか，熟練した他者の協力を得るなどしないとわからないという問題が残る．

一方，2.3.1 項で述べたように，特定のタスクやシーンにおける言語・非言語行動から，個人の性格特性やコミュニケーションスキル，共感スキルといった能力やパフォーマンスを推定する取り組みも行われている [30,37]．これらの研究事例は，褒める行為に関連しうるコミュニケーションスキルや共感スキルを扱ったものの，褒める行為を直接的に扱ったものではない．そのため，コミュニケーション中の言語・非言語行動から，褒める行為を検出することができるのか，褒め方の上手さを推定することができるのか，明らかにされているとはいえない．

3.1.2　理解を表す行為に関する既存研究の問題点

コミュニケーションにおいて，聞き手が話し手の話に対して，理解しているか否かを示すことは重要である．一方で，話し手が聞き手が自身の話を理解しているか推測することも重要である．これらの行為がコミュニケーション中に行われることで，コミュニケーションが円滑に進むことが期待される．しかし，コミュニケーションが苦手な人は，自身が理解できている，もしくは理解できていないことを対話相手に伝えられていない可能性がある．さらに，そのような人は，自身の発言に対して対話相手が理解できている，もしくは理解できていないことを把握できていない可能性がある．

2.3.2.2 目で述べたように，医療現場，ポスター発表などの特定シーンにおける理解度を推定する取り組みが行われている [90–92]．これらの研究より，医療行為の説明やポスター発表といったコミュニケーションシーンに限定した人々の理解度はある程度推定できることが明らかになっている．しかし，医療行為の説明やポスター発表のような専門的な知識を必要としない，日常生活で行われるコミュニケーションに近いシーンにおける理解度を推定する研究は行われていない．そのため，コミュニケーション中の聞き手の理解度を聞き手の言語・非言語行動から推定できるのかは明らかになっていない．

3.1.3　相槌を打つ行為に関する既存研究の問題点

聞き手の相槌は，対話を成立させ，円滑に進めるために重要である．日本語対話における相槌は，継続を促す表現，理解を示す表現，同意をする表現，感情の表現，情報を要求する表現などの機能別に分類することができる [39]．しかし，2.3.2.1 目で述べた相槌を推定する多くの既存研究では，相槌の有無やタイミングを推定する取り組みが多く [70–78] 相槌の機能を推定する取り組みは限定的である [50, 80]．上記の取り組みから，コミュニケーション中に相槌を打つ適切なタイミングは明らかになるが，打つべき相槌の機能が適切であるのか不明である．そのため，コミュニケーション中に聞き手が打つべき相槌が，コミュニケーションの流れや内容に適しているのか話し手の言語・非言語行動から判断することができないという問題がある．

3.2　研究課題の設定

本研究では，話し手および聞き手の言語・非言語行動から聞き手のフィードバックに対する評価を推定する機械学習モデルや，聞き手のフィードバックの種類を分類する機械学習モデルを構築する取り組みを行う．聞き手のフィードバックは，継続を表す表現，内容理解を示す表現，話し手の判断を支持する表現，相手の意見や考えに肯定・同意する表現，感情の表現，情報の追加・訂正・要求をする表現，話題転換を促す表現に分類できる [39, 93, 94]．これらの中で本研究では，話し手と聞き手の両者に対して好意的な相互作用を及ぼす聞き手が内容理解を示す表現，相手の意見や考えに肯定・同意する表現，感情の表現に着目する．具体的には，より良好な人間関係の構築が期待でき，より円滑なコ

ミュニケーションを実現する上で重要な役割を果たす (1) 褒める行為 [40, 41]，(2) 理解を表す行為 [8, 9, 16]，(3) 相槌を打つ行為 [9, 39] に着目し，次の 3 つの課題に取り組む．

第 1 の課題では，話し手および聞き手の言語・非言語行動から褒める行為の検出や褒め方の上手さを推定する取り組みを行う [51–53]．具体的には，実対話データと褒め方に関する評価を記録した対話コーパスを構築し，視覚的，韻律的，言語的特徴量を用いて褒める行為を検出する機械学習モデル，および褒め方の上手さを推定する機械学習モデルを構築する取り組みを行う．

第 2 の課題では，聞き手の言語・非言語行動から聞き手の理解度を推定する取り組みを行う [54]．具体的には，実対話データと聞き手の理解度を記録した対話コーパスを利用し，聞き手の視覚的，韻律的，言語的特徴量から聞き手の理解度を推定する機械学習モデルを構築する取り組みを行う．

第 3 の課題では，話し手の言語・非言語行動から聞き手の相槌機能を予測する取り組みを行う [55]．具体的には，実対話データと聞き手の相槌機能を記録した対話コーパスを利用し，視覚的，韻律的，言語的特徴量から聞き手の相槌機能を分類する機械学習モデルを構築する取り組みを行う．

第4章　聞き手の褒める行為を自動検出・褒め方の上手さを自動推定する取り組み

4.1　聞き手の褒める行為の自動検出・褒め方の上手さの自動推定を行う重要性

日常生活や社会活動において，褒める行為は重要なコミュニケーション方法のひとつである．褒める行為は，教育学や心理学などの学術分野で次のように定義されている．

- 対象の行動に向けて肯定的なフィードバックを行うことや，望ましい行動に対して承認や賞賛を表す言葉やジェスチャであること [95].
- 対象の行動や性格に向けた承認や賞賛を表す言語･非言語行動であるとこと [40,46,47].

このことから，褒める行為は対象の行動に対して承認や賞賛を表す言語・非言語行動であると考えられる．さらに，褒める行為は，褒める人から褒められる人への一方的な意思の伝達ではなく，褒められる人の役割は褒める人の役割と同じくらい重要であると考えられている [41,96]．このことから，褒める人の行動だけてなく，褒められる人の行動にも注目する必要があると考えられる．

教育学や心理学などの学術分野において，褒める行為に関する研究は多く行われている [41,63–69]．教育や育児におけるシーンで，教師が生徒を褒めることや保護者が子供を褒めることが生徒や子供の行動，モチベーション，感情に影響するのか調査する研究が多い．その結果，生徒や子供は肯定的なフィードバックを与えられると，その後の行動を続ける時間が長くなることや，ポジティブな表情やジェスチャが表出することが報告されている [68,69]．しかし，多くの研究では，褒める行為が対象に及ぼす影響を調査しており，自身が褒めることができているのか，その褒め方はどれくらい上手いのか把握するための研究は調査する限り行われていない．そのため，人とのコミュニケーションが苦手な人やコミュニケーションスキルが乏しい人は，熟練した他者の協力を得るなどしないと対話相手を褒めることができているのか，どれくらい上手く褒めることができているのかを把握することができないという問題がある．

この問題を解決するために，聞き手は褒める行為を行っているのか，どれくらい上手く褒めているのか自動的に判定する必要がある．自動判定を実現することで，コミュニケーションシーンやすでに持っている個々の知識や技能に合わせた褒め方の上手さを評価できるようになると考えられる．上記の実現を目指し，本研究は，話し手および聞き手の言語・非言語行動から褒める行為の検出や褒め方の上手さを推定する取り組みを行う [51–53].

4.2 節では，聞き手の褒める行為を検出・褒め方の上手さを推定するための対話コーパスについて説明する．4.3 節では，聞き手の褒める行為を検出する機械学習モデルについて説明する．4.4 節では，聞き手の褒め方の上手さを推定する機械学習モデルについて説明する．最後に，4.5 節では，聞き手の褒める行為を検出・褒め方の上手さを推定する取り組みについて総括する．

4.2　褒める行為の検出・褒め方の上手さの推定を行うための対話コーパス

本節では，対話中の褒める行為を分析するための対話コーパス [97] について述べる．この対話コーパスには，2 者対話における話者の言語・非言語行動と褒める行為に関する評価値が含まれている*．本章では，対話における話し手を受け手（褒められる人），聞き手を褒め手（褒める人）とする．

4.2.1　2 者対話データ

本項では，2 者対話データについて説明する．この 2 者対話データには，日本人同士による対面の 2 者対話が記録されている．

4.2.1.1　対話の参加者について

2 者対話の参加者は，20 代の大学生 34 名（男性 28 名，女性 6 名）であり，母国語が日本語で，文系・理系を問わず様々な学問分野を専攻している．対話参加者は，2 名 1 組のペア（計 17 組）に分けられた．このとき，可能な限り初対面同士のペアとなるように，実験者が対話参加者の所属学科を確認した上で対話参加者のペアを組み合わせた．その結果，17 組のうち，初対面が 14 組，顔見知りが 2 組，友人同士が 1 組であった．

4.2.1.2　対話の収録環境について

2 者対話の収録は，大学構内の静かな部屋で実施した（図 4.1）．この部屋には，実験者と参加者以外は立入らないようにした．さらに参加者が対話を行う区画と，実験者が待機する区画をパーテーションで区切るようにした．対話参加者は，互いに向き合って着座して対話を行った．このとき，参加者間の距離は 180cm とした．

対話は，各参加者の様子と 2 者対話全体の様子を撮影するためのビデオカメラと各参加者の音声を録音するためのマイクで記録した．ビデオカメラは SONY FDR-X3000 Action Cams を用い，解像度が 1980×1080，ビットレートが 16 Mbps の映像データを記録した．マイクは JTS CM-214i のヘッドセットを用い，Roland R-44 のレコーダを経由してサンプリングレートが 44.1 kHz の音声データを記録した．

4.2.1.3　対話の収録手順について

各組の参加者ら（参加者 A と参加者 B）は，実験者の指示に従い，対話を行った．対話の収録前に，参加者は過去に頑張ったことに関するエピソードを 2 つ以上用意した．これ

*対話コーパスの構築にあたり，日本大学文理学部研究倫理委員会の承認を得ている．

図 4.1: 2 者対話収録の様子

は，対話が始まってからエピソードを考えると，対話が円滑に進まない可能性があるためである．エピソードは，受験，サークル活動，日常生活，趣味に関するものが多かった．対話は，次の流れで行った．はじめに，互いに自己紹介を行う対話を 5 分間行った．この対話は，多くのペアが初対面同士であるため，緊張をほぐすことを目的としている．次に，参加者 A を褒め手（褒める人），参加者 B を受け手（褒められる人）となる対話を 5 分間行った．最後に，各参加者の役割を入れ替え，参加者 B を褒め手（褒める人），参加者 A を受け手（褒められる人）となる対話を 5 分間行った．本対話コーパスでは，上記の対話を 17 組分，計 255 分間のデータを記録している．

本研究では褒め手（褒める人）を対話相手に対して褒める行為を行う人と定義した．対話を行う際に，褒め手には，対話相手を積極的に褒めるように指示した．しかし，一方的に褒めているだけのような不自然な対話にならないようにするために，自由に質問したり，リアクションをしたりすることを許可した．

一方で，本研究では受け手（褒められる人）を対話相手から褒める行為を受ける人と定義した．受け手には，事前に用意した自分がいままで頑張ってきたことに関連するエピソードを話すように指示した．加えて，対話の自然さや話題の多様性を担保するために，事前に用意していないエピソードについて話すことを許可した．

4.2.2　アノテーション

本研究では，発話区間および発話内容に関するアノテーションと褒める行為に関するアノテーションを行った．4.2.2.1 目では，発話区間および発話内容のアノテーションについて説明する．4.2.2.2 目では，褒める行為に関するアノテーションについて説明する．

4.2.2.1　発話区間および発話内容のアノテーション

映像や音声データに対して注釈付けを行うツールである ELAN [98] を用いて，対話に参加していないアノテータが，各対話参加者の音声データを参照して，発話区間の抽出と発話内容の書き起こしを行った．発話区間は，無音区間が 400ms 未満の連続した音声区間（Inter-pausal units, IPU [99]）とした．多くの研究では，IPU の区切りを 200ms [†]としているが，本研究の対話データでは発話が細かく区切れてしまうため，自然な区切りとなるように 400ms [100] とした．この結果，褒め手の発話区間が 2,701 件，受け手の発話区間が 3,413 件抽出された．

発話内容は，アノテータによって音声データから書き起こされた．書き起こしを行う際，アノテータによる書き起こし内容の表記ゆれが分析結果に影響を及ぼすことが考えられた．表記ゆれを減らすために，アノテータは辞書を参照しながら作業を行った．本研究は，UniDic[‡]の話し言葉解析用辞書 [101] に載っている感動詞とフィラーの単語を抜き出したものを辞書として扱った．このときアノテータには，発話内容の文頭が辞書内の単語の読みに近い場合は，辞書内の単語で書き起こすように指示した．たとえば，“あー”，“あ〜” などは “あー” と書き起こされる．なお，辞書に載っていない単語は，聞き取った単語をそのまま書き起こされている．

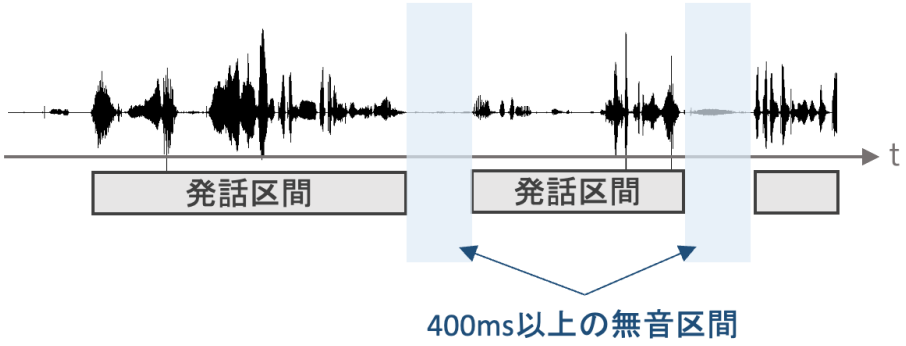

図 4.2: 発話区間の一例

4.2.2.2　褒める行為に関するアノテーション

本研究が属する社会的信号処理の分野では，人間による能力やパフォーマンスのアノテーションが広く行われている [79, 102]．主観的な判断が要求されるアノテーションタスクを 1 名のアノテータが実施する場合，偏りやラベル付けの揺れが生じる可能性があ

[†]国立国語研究所「日本語話し言葉コーパス」において，転記基本単位（言語音が 0.2 秒以上の途切れがなく，連続して生じている区間）と定義されている．

[‡]国立国語研究所の規定した斉一な言語単位と，階層的見出し構造に基づく電子化辞書

る．複数のアノテータが担当することで，バイアスの可能性を低減することが期待できる．さらに，多くの研究では，話者自身ではなく，第三者がアノテーションを行っている [103–107].

そこで本研究では，4.2.1 項の 2 者対話収録と 4.2.2.1 目の発話に関するアノテーションに参加していないアノテータ 5 名が，褒める行為に関するアノテーションを行った．アノテータは，予備知識や先入観の影響を避けるため，訓練や資格の取得などは行わなかった．アノテータは，褒め手の正面に設置したビデオカメラで撮影した映像データと，褒め手に取り付けたマイクで録音した音声データを視聴し，褒め手の発話区間ごとに次の判定と評価を行った．

- 対話相手を褒めているシーンであるか，そうでないかの判定
- 褒めているシーンであると判定した場合，褒め方の上手さを 1（上手く褒めていない）〜7（上手く褒めている）の 7 段階で評価

1 つ目の判定において，アノテータ間の判定の一致度を確認するために，Fleiss のカッパ係数 [108] を利用して，アノテータの判定の一致率を算出した．その結果，カッパ係数は 0.698 であり，アノテータ間の判定の一致度は高いことがわかった．さらに，2 つ目の評価において，アノテータ間の評価の一致度を確認するために，級内相関係数（Intraclass correlation coefficients, ICC）[109] を利用して，アノテータの評価の一致率を算出した．一致率の算出には，褒めているシーンであると判定したアノテータが過半数に達したシーンにおいて，褒めているシーンであると判定したアノテータの褒め方の上手さの評価結果を用いた．具体的には，当該シーンにおいて褒めているシーンであると判定したアノテータの人数の組み合わせごとに級内相関係数を算出した．その後，各組み合わせのサンプル数を考慮して重み付き平均を算出した．その結果，$ICC(2,k) = 0.571$ であり，アノテータ間の評価の一致度は中程度であることがわかった [110].

上記のとおり，本研究では，第三者による判定と評価を行った．褒める行為は，実際に褒められた人と，客観的に褒めている様子を確認している第三者で，印象が異なる可能性が存在する．そこで，褒められた本人とアノテータの評価に違いがあるか確認するために，対話参加者にも褒める行為に関する判定と評価を行ってもらった．具体的には，対話参加者 5 名に自身が受け手であった対話における褒め手の映像データと音声データを視聴し，褒め手の発話区間ごとに次の判定と評価を行った．

- 対話相手を褒めているシーンであるか，そうでないかの判定
- 褒めているシーンであると判定した場合，褒め方の上手さを 1（上手く褒めていない）〜7（上手く褒めている）の 7 段階で評価

対話参加者 5 名が判定，評価した結果と第三者のアノテータ 5 名が判定，評価した結果の相関係数を確認した，その結果，相関係数が 0.773 であり，対話参加者が評価した場合でも，アノテータが評価した場合でも，褒める人の褒め方に対する印象におおむね違いがないことが確認できた．

4.2.2.3　Praise シーンと Praise スコアの定義

4.2.2.2 目で判定したアノテーション結果に基づき，Praise シーン，Non-praise シーンを定義する．

Praise シーン： 対話相手を褒めているシーンであると判定したアノテータが 3 名以上の発話区間

Non-praise シーン： 対話相手を褒めているシーンであると判定したアノテータが 2 名以下の発話区間

上記の定義に基づくと，本研究の対話コーパスでは，Praise シーンが 228 件，Non-praise シーンが 2,473 件含まれていた．

次に，4.2.2.2 目で評価したアノテーション結果に基づき，Praise スコアを定義する．

Praise スコア： 各 Praise シーンにおいて，対話相手を褒めているシーンであると判定したアノテータの褒め方の上手さの評価の平均値

Praise スコアの分布を図 4.3 に示す．

最後に，Praise スコアに基づいて，Praise シーンを 3 つの群に分割する[§]．

Praise スコア低群： Praise スコアが 3.8 点未満の Praise シーン（計 82 シーン）

Praise スコア中群： Praise スコアが 3.8 点以上 4.4 点未満の Praise シーン（計 65 シーン）

Praise スコア高群： Praise スコアが 4.4 点以上の Praise シーン（計 81 シーン）

4.3　褒める行為を検出する機械学習モデルの構築

本節では，褒め手および受け手の言語・非言語行動から褒める行為を検出する機械学習モデルについて述べる．具体的には，視覚的，韻律的，言語的特徴量を用いて褒める行為を検出する機械学習モデルを構築する取り組みを行う．

4.3.1　特徴量抽出

本研究は，4.2.1 項で記録した対話データから褒め手と受け手の視覚的，韻律的，言語的特徴量を抽出した．

具体的には，4.2.2.3 目で定義した Praise シーンとその前後 1 秒間を加えた範囲を特徴量の抽出範囲と設定し，その範囲における褒め手と受け手の視覚的，韻律的，言語的特徴量を抽出した．

[§]各群に属するシーン数は等しいことが理想ではあるが，スコアが同じシーンが多数存在したため，各群のシーン数を等しくすることができなかった．

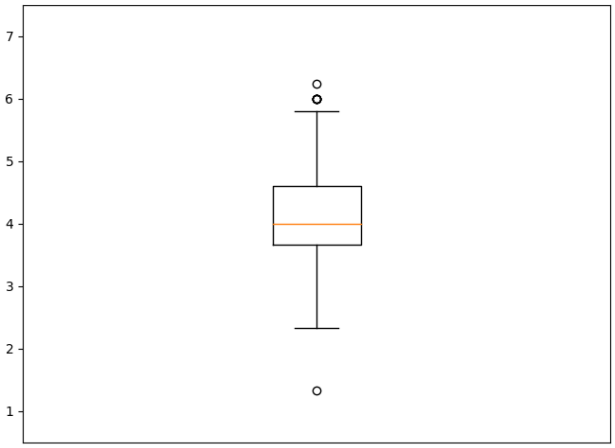

図 4.3: Praise スコアの分布（N=228）

4.3.1.1　視覚的特徴量

本研究は，視覚的特徴量として頭部，視線，表情に関する行動に着目する．頭部は，例示的動作，調整的動作，情動表出，適応動作に分類できるとされており，コミュニケーションにおいて言語行動の補完や相手への合図（肯定や否定の動作や傾聴の動作）の機能を果たすと考えられている [111]．視線は，身体行動には表れない感情，態度，関係に関する信号を提供すると考えられており，コミュニケーションにおいて多くの機能を果たすと考えられている [15]．表情は，人間の感情と密接に関係しており，自分自身，相手，他者に対する感情に関する豊富な情報を提供する手がかりであると考えられている [4]．

本研究は，顔画像処理ツールである OpenFace [112] を用いて，褒め手（praiser_）と受け手（receiver_）の正面に設置したビデオカメラで撮影した映像データから頭部，視線，表情（Action Units [84]）に関する特徴量を抽出した．これらの特徴量は，既存研究 [29] を参考にしている．抽出した視覚的特徴量は，合計 88 次元である．特徴量の詳細は次のとおりである．

頭部： ビデオカメラ側から顔を見て，左から右方向を x 軸，下から上方向を y 軸，手前から奥方向を z 軸とした場合，Praise シーンの前後 1 秒ずつを含む範囲における頭部の x 軸，y 軸，z 軸周りの回転角度（pose_Rx，pose_Ry，pose_Rz) の分散（ _var），中央値（_med），10 パーセンタイル値（_p10），90 パーセンタイル値（_p90）を用いた．

視線： ビデオカメラ側から顔を見て，左から右方向を x 軸，下から上方向を y 軸とした場合，Praise シーンの前後 1 秒ずつを含む範囲における視線の x 軸，y 軸方向の角度

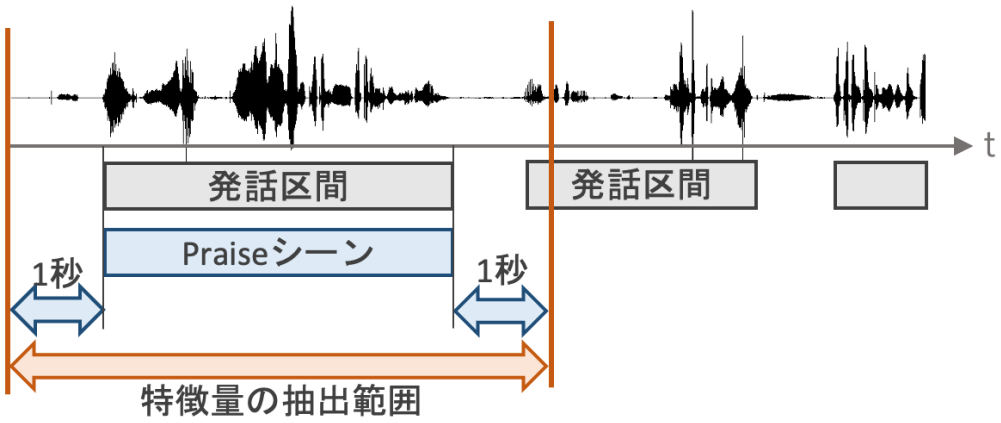

図 4.4: 褒める行為の検出・上手さの推定における特徴量抽出範囲

（gaze_Ax，gaze_Ay）の分散，中央値，10 パーセンタイル値，90 パーセンタイル値を用いた．

Action Units： Action Units は，人間の顔部における筋肉群の基本的な行動の単位を表している [84]．Praise シーンの前後 1 秒ずつを含む範囲における OpenFace で用いられている各 Action Units(表 4.1) の強度の分散，中央値，10 パーセンタイル値，90 パーセンタイル値を用いた．

表 4.1: OpenFace での Action Units の内容

項目	内容	項目	内容
AU01	眉の内側を上げる	AU14	笑窪を作る
AU02	眉の外側を上げる	AU15	唇の両端を下げる
AU04	眉を下げる	AU17	顎を上げる
AU05	上瞼を上げる	AU20	唇の両端を横に引く
AU06	頬を持ち上げる	AU23	唇を固く閉じる
AU07	瞳を緊張させる	AU25	顎を下げずに唇を開く
AU09	鼻に皺を寄せる	AU26	顎を下げて唇を開く
AU10	上唇を上げる	AU45	瞬きをする
AU12	唇の両端を引き上げる		

4.3.1.2　韻律的特徴量

本研究は，韻律的特徴量として発話にともなう音声手がかりである声質 [113] に着目する．音声は，感情，健康状態，年齢，性別などの情報を含んでおり，テンポ，ピッチなどの声質を変化させることで言語情報（テキスト）では伝えられない情報を伝えることができると考えられている [58].

本研究は，音声処理ツールである openSMILE [114] を用いて，褒め手と受け手に取り付けたマイクで録音した音声データから，韻律の代表的な特徴量と発話時間に関する特徴量を抽出した．韻律の代表的な特徴量について，本研究では表 4.2 に示す韻律の代表的な特徴量ごとに，表 4.3 に示す統計量を算出した．さらに，上記の特徴量に対して一次微分した特徴量（_de）を抽出した．これらの特徴量は，openSMILE の標準セット [115] として提供されたものである．発話時間に関する特徴量は，褒め手の発話の長さ（秒）とした．抽出した韻律的特徴量は，合計 384 次元であった．

表 4.2: 韻律の代表的な特徴量

特徴量	内容
pcm_RMSenergy_sma	音のエネルギーの二乗平均平方根
pcm_fftMag_mfcc_sma[1]–[12]	1〜12 次のメル周波数ケプストラム係数
pcm_zcr_sma	ゼロ交差率
voiceProb_sma	声である確率
F0_sma	基本周波数

表 4.3: 抽出に用いた統計量

統計量	内容	統計量	内容
_max	最大値	_linregc1	線形近似の勾配
_min	最小値	_linregc2	線形近似のオフセット
_range	最大値と最小値の差	_linregerrQ	線形近似の二乗誤差
_maxPos	最大値の絶対位置	_stddev	標準偏差
_minPos	最小値の絶対値位置	_skewness	歪度
_amean	平均値	_kurtosis	尖度

4.3.1.3　言語的特徴量

本研究は，言語的特徴量として発話内容の品詞と発話内容のベクトル表現に着目する．本研究は，4.2.2.1 目で書き起こされた発話内容から，品詞に基づく言語的特徴量（以降，品詞特徴量）および発話内容をベクトル化した言語的特徴量（以降，発話埋め込み特徴量）を用いて，言語行動に関連する特徴量を抽出した．

品詞： 品詞特徴量は，当該シーンにおける発話内容の各品詞の出現頻度と単語数からなる．本研究では，日本語話者の発話内容を対象としたため，日本語における品詞を用いた．日本語の形態素解析システムである MeCab [116] を用いて，各発話区間の褒め手の発話内容に形態素解析を行い．各単語の品詞を抽出した．抽出した品詞を表 4.4 に示す．単語数は，各発話区間における各品詞の出現数を集計したものとした．品詞の出現頻度は，各発話区間における各品詞の出現数をその発話区間での単語数で割ったものとした．品詞特徴量は，全部で 37 次元であった．

発話埋め込み： 発話埋め込み特徴量は，当該シーンにおける発話内容の 768 次元のベクトル（文書埋め込み）からなる．BERT は，大規模コーパスを用いて事前学習したモデルを転移学習により様々なタスクに応用できる言語モデルである [117]．具体的な抽出方法は，事前に学習した日本語言語モデルである BERT モデル[¶]を用いて，入力した文章を BERT が処理した最終層の先頭にある特殊トークン（CLS[‖]）の埋め込み表現を取得した．この方法は，既存研究 [38] を参考にした．発話埋め込み特徴量は合計 768 次元であった．

4.3.2　機械学習モデルの構築

本項では，言語・非言語行動に関する特徴量から褒める行為を検出する機械学習モデルについて述べる．本研究は，目的変数を 4.2.2.3 目で定義した Praise シーン，Non-praise シーンとし，説明変数を 4.3.1 節で抽出した特徴量とする機械学習モデルを構築した．機械学習モデルの構築では，アンサンブル学習のバギング手法であり，少ないデータ数の場合でも性能が良くなることが見込まれる Random Forests [118] を用いた．本研究が属する分野において，Random Forests を用いている研究事例 [32] は少ないが，各特徴量の重要度を算出できるため，フィードバックの評価や種類の判断基準となる言語・非言語行動が明らかになると考えられる．ハイパーパラメータは，Hyperopt [119] を用いて，木の本数と木の深さをチューニングした．各パラメータの探索範囲は次のとおりである．

木の本数（n_estimators）： 10〜100

木の深さ（max_depth）： 3〜10

[¶] https://github.com/cl-tohoku/bert-japanese
[‖] 文頭に埋め込むトークンを表す

表 4.4: 抽出した品詞

大分類	小分類
フィラー	フィラー
感動詞	感動詞
形容詞	形容詞
助詞	助詞（格助詞），助詞（副助詞），助詞（連体化）， 助詞（接続助詞），助詞（特殊），助詞（副詞化）， 助詞（副助詞），助詞（副助詞／並立助詞／終助詞）， 助詞（並立助詞），助詞（連体化），
助動詞	助動詞
接続詞	接続詞
動詞	動詞（自立），動詞（接尾），動詞（非自立）
名詞	名詞（サ変接続），名詞（ナイ形容詞語幹），名詞（一般）， 名詞（引用文字列），名詞（形容動詞語幹），名詞（固有名詞）， 名詞（数），名詞（接続詞的），名詞（接尾），名詞（代名詞）， 名詞（動詞非自立的），名詞（特殊），名詞（非自立）， 名詞（副詞可能）
連体詞	連体詞
副詞	副詞
接頭詞	接頭詞
記号	記号

ハイパーパラメータのチューニングは，訓練データ 90%とテストデータ 10%に分割されたデータを用いて，1,000 回行った．1,000 回の試行結果に基づいて，最も性能の良いパラメータを用いることとした．さらに，モデルの性能の向上が止まるまで，特徴量重要度（Gain）の低い特徴量を削除する方法で特徴量を選択した．Random Forests における特徴量重要度は，各特徴量が不純度の減少に寄与した量の合計を正規化した値であり，各特徴量の寄与度を示している [118]．その中で，Gain は各特徴量が Random Forests 内の決定木で分割に寄与した不純度の減少量を累積し，正規化した値である．具体的には，次のように算出される**．はじめに，各決定木のノードで親ノードと子ノードの不純度の差を算出する．このとき，算出した差を減少量とする．次に，Random Forests 全体で，各決定木の分割で用いられた特徴量ごとに算出された減少量を合計する．最後に，全特徴量の減少量の合計が 1 となるように正規化する．

本研究では，次のタスクを 100 回繰り返して行った．

(1) データ数を揃えるために，Non-praise シーンを無作為に 228 件選択する．

**本研究は，scikit-learn を用いて特徴量重要度を算出した（https://scikit-learn.org/dev/auto_examples/ensemble/plot_forest_importances.html）

(2) データセットを，訓練データ 90%，テストデータ 10%に無作為に分ける．このとき，訓練データとテストデータにおいて各クラスに属するシーン数の割合が同じになるように分割した．

(3) 訓練データで構築したモデルを用いて，テストデータが属するクラスを分類する．

私の知る限り，既存の研究では褒める行為を定量化する方法は特定されておらず，したがって適切なベースライン法が提供されていない．比較のための既存の手法が利用できないことから，本研究ではベースライン手法として，データセット内のデータ分布に基づいて，対象シーンの行為がどのグループに属するかを確率的に決定するアプローチを採用した．具体的には，ベースラインモデルは，50%の確率で Praise シーン，Non-praise シーンを検出するモデルを採用した．

4.3.3　機械学習モデルの性能

機械学習モデルの性能を表 4.5 に示す．モデルの性能を評価するにあたり，F 値の重み付き平均を用いた．本研究が属する分野において，分類タスクを採用している研究事例 [72–74, 78, 91, 120] は再現率，適合率，F 値（F1-score）を用いている．F 値は，再現率（recall）と適合率（precision）の調和平均である．再現率（recall）は，機械学習モデルが真であると予測した数を分母に，正しく予測できていた数を分子にした値である．適合率（precision）は，正解データの真である数を分母に，モデルが正しく予測できていた数を分子にした値である．これらの式を次に示す．このとき，正しく予測できた値を TP，真であるのに偽と予測した値を FN，偽であるのに真と予測した値を FP とする．

$$\text{再現率} = \frac{TP}{TP + FN} \tag{4.1}$$

$$\text{適合率} = \frac{TP}{TP + FP} \tag{4.2}$$

$$\text{F 値} = \frac{2 \cdot recall \cdot precision}{recall + precision} \tag{4.3}$$

本研究は，算出した F 値に対して重み付き平均を算出した．具体的には，次の式で算出した．このとき，各クラスの F 値を F_i，各クラスのデータ数を n_i クラス数を m とする．

$$\text{F 値の重み付き平均} = \frac{\sum_{i=0}^{m} F_i n_i}{\sum_{i=0}^{m} n_i} \tag{4.4}$$

表 4.5 の F 値は，各モデルを 100 回構築した際の重み付き F 値の平均値である．表 4.5 に示すように，ベースラインの検出性能（F 値）は 0.502 であった．本稿で提案した機械学習モデルの中で，最も検出性能が高かったモデルは M55 であり，検出性能（F 値）は 0.827 であった．M55 で得られたパラメータは次のとおりであった．

木の本数（n_estimators）： 64

木の深さ（max_depth）： 6

表 4.5: 褒める行為を検出する機械学習モデルの性能

	褒め手				受け手				褒め手				受け手		
	非言語		言語		非言語		F 値		非言語		言語		非言語		F 値
	視覚的	韻律的	言語的（品詞）	言語的（BERT）	視覚的	韻律的			視覚的	韻律的	言語的（品詞）	言語的（BERT）	視覚的	韻律的	
Baseline							0.502	M32	✓			✓		✓	0.811
M1	✓						0.654	M33	✓				✓	✓	0.624
M2		✓					0.761	M34		✓	✓		✓		0.758
M3			✓				0.595	M35		✓	✓			✓	0.745
M4				✓			0.822	M36		✓		✓	✓		0.822
M5					✓		0.567	M37		✓		✓		✓	0.807
M6						✓	0.533	M38		✓			✓	✓	0.759
M7	✓	✓					0.758	M39			✓	✓	✓		0.803
M8	✓		✓				0.658	M40			✓	✓		✓	0.798
M9	✓			✓			0.803	M41			✓		✓	✓	0.575
M10		✓	✓				0.758	M42				✓	✓	✓	0.813
M11		✓		✓			0.810	M43	✓	✓	✓		✓		0.763
M12			✓	✓			0.806	M44	✓	✓	✓			✓	0.777
M13	✓	✓	✓				0.766	M45	✓	✓		✓	✓		0.800
M14	✓	✓		✓			0.810	M46	✓	✓		✓		✓	0.802
M15	✓		✓	✓			0.808	M47	✓	✓			✓	✓	0.775
M16		✓	✓	✓			0.801	M48	✓		✓	✓	✓		0.807
M17	✓	✓	✓	✓			0.818	M49	✓		✓	✓		✓	0.806
M18					✓	✓	0.566	M50	✓		✓		✓	✓	0.649
M19	✓				✓		0.652	M51	✓			✓	✓	✓	0.787
M20	✓					✓	0.620	M52		✓	✓	✓	✓		0.807
M21		✓			✓		0.759	M53		✓	✓	✓		✓	0.823
M22		✓				✓	0.756	M54		✓	✓		✓	✓	0.752
M23			✓		✓		0.632	M55		✓		✓	✓	✓	**0.827**
M24			✓			✓	0.548	M56			✓	✓	✓	✓	0.797
M25				✓	✓		0.816	M57	✓	✓	✓	✓	✓		0.804
M26				✓		✓	0.817	M58	✓	✓	✓	✓		✓	0.808
M27	✓	✓			✓		0.753	M59	✓	✓	✓		✓	✓	0.770
M28	✓	✓				✓	0.749	M60	✓	✓		✓	✓	✓	0.810
M29	✓		✓		✓		0.658	M61	✓		✓	✓	✓	✓	0.819
M30	✓		✓			✓	0.638	M62		✓	✓	✓	✓	✓	0.825
M31	✓			✓	✓		0.807	M63	✓	✓	✓	✓	✓	✓	0.808

4.3.4　褒める行為の検出に寄与するモダリティについて

本項では，褒める行為の検出に寄与したモダリティについて述べる．

4.3.4.1　ユニモーダル特徴量を用いた検出

ユニモーダル特徴量を用いて褒める行為を検出した場合（M1-M6），ベースラインモデルよりも検出性能を向上させることができた．このことから，本研究で利用した褒め手のすべての特徴量と受け手のすべての特徴量が検出に有用であると考えられる．その中で，最も高い検出性能を達成したのは，褒め手の発話埋め込み特徴量を用いたモデル（M4）であり，検出性能（F 値）は $F = 0.822$ であった．今回構築したモデルの中では，褒め手の発話埋め込み特徴量を用いたモデルが 32 件，褒め手の発話埋め込み特徴量を用いていないモデル 31 件ある．褒め手の発話埋め込み特徴量を用いたモデルが，検出性能上位 32 件を占める結果となった．このことから，褒め手の発話埋め込み特徴量が褒める行為を検出する上で重要な特徴量であると考えられる．

4.3.4.2　褒め手の特徴量を用いた検出

褒め手の特徴量を用いて褒める行為を検出した場合（M1-M4, M7-M17），ベースラインモデルよりも検出性能を向上させることができた．その中で，最も高い検出性能を達成したのは，褒め手の発話埋め込み特徴量を用いたモデル（M4）であり，検出性能（F 値）は $F = 0.822$ であった．このモデルは，褒め手のすべての特徴量を用いたモデル（M17）よりも検出性能が高かった．そのため，褒め手のすべてのモダリティを用いなくても褒める行為を十分に検出できることが示唆された．

4.3.4.3　受け手の特徴量を用いた検出

受け手の特徴量を用いて褒める行為を検出した場合（M5，M6，M18），ベースラインよりも検出性能を向上させることができた．その中で，最も高い検出性能を達成したのは，受け手の視覚的特徴量を用いたモデル（M5）であり，検出性能（F 値）は $F = 0.567$ であった．このモデルは，ベースラインよりも検出性能が高かったものの，褒め手の特徴量を用いたモデルよりも検出性能が高くないため，受け手のみでは褒める行為を検出することは難しいと考えられる．

4.3.4.4　褒め手と受け手の特徴量を用いた検出

両者（褒め手と受け手）の特徴量を用いて褒める行為を検出した場合（M19-M63），ベースラインモデルよりも推定性能を向上させることができた．その中で，最も高い検出性能を達成したモデルは，褒め手の韻律的特徴量と発話埋め込み特徴量と受け手の視覚的特徴量と韻律的特徴量を用いたモデル（M55）であり，検出性能（F 値）が $F = 0.827$ であった．このモデルは，4.3.2 項で構築したモデルの中で，最も検出性能が高かった．このことから，褒め手と受け手の両者の特徴量を用いることが褒める行為の検出に有用であると考えられる．その中で，褒め手の韻律的特徴量と発話埋め込み特徴量と受け手の視覚的特徴量と韻律的特徴量に関するモダリティが，対話中の褒める行為を検出する上で重要であると考えられる．

4.3.4.5　ユニモーダル特徴量で検出した場合とマルチモーダル特徴量で検出した場合の比較

ユニモーダル特徴量で最も検出性能が高かった M4（F=0.822）と，マルチモーダル特徴量で最も検出性能が高かった M55（F=0.827）の間で，対応のない t 検定（Welch の t 検定，有意水準 5%）を行った．その結果，M4 と M55 の間に統計的に有意な差は認められなかった．100 回の検出結果において，M55 は M4 に比べて F 値が向上した．これは，M4 がユニモーダルを用いたことに対し，M55 はマルチモーダルな特徴量を用いたことに起因すると考えられる．具体的には，M55 は褒め手の韻律的特徴量と発話埋め込み特徴量と受け手の視覚的特徴量と韻律的特徴量を用いている．対照的に，M4 は褒め手の発話

埋め込み特徴量のみを用いた．既存研究において，マルチモーダルモデルの性能がユニモーダルモデルを上回ることが報告されており [29, 32]，異なるモダリティが相乗的に作用していることが影響していると考えられる [29]．

4.3.4.6　褒める行為の検出における特徴量重要度について

褒める行為の検出に寄与したモダリティを明らかにするために，特徴量重要度（Gain）を算出した．最も検出性能の高かったモデル（M55）における特徴量重要度を図 4.5 に示す．図 4.5 より，褒める行為の検出に寄与していた特徴量上位 10 件は，褒め手のゼロ交差率，メル周波数ケプストラム係数（MFCC），発話埋め込みに関する特徴量であった．ゼロ交差率は，音の波形が音圧 0 である軸を交差する割合を表す特徴量である．MFCC は，対数ケプストラム（20 次）の低次成分（1〜12 次）に対して人間の周波数知覚特性を考慮して重み付けをしたものであり，声道特性を表す特徴量である．このことから，M55 で用いた特徴量（褒め手の韻律的，発話埋め込み特徴量と受け手の視覚的，韻律的特徴量）の中で，褒め手のゼロ交差率，メル周波数ケプストラム係数（MFCC），発話埋め込みに関する特徴量が褒める行為の検出に寄与していたことが確認された．これらの特徴量の分布を確認するために，Praise シーンと Non-praise シーンにおける各特徴量の平均値と標準偏差を表 4.6††に示す．Praise シーンと Non-praise シーンの間に大小関係があるのか確認するために，対応のない t 検定を行った．検定の結果，褒める行為の検出に寄与していた重要度上位 10 件の特徴量は，Praise シーンと Non-praise シーンの間に 5%水準の有意差が認められた．このことから，重要度上位 10 件の特徴量は，Praise シーンと Non-praise シーンの間に大小関係があると考えられる．表 4.6 より，ゼロ交差率に関する特徴量の値は，Praise シーンの方が大きくなることが確認できる．このことから，褒め手が褒める発話を行う際の音声は，音圧 0 である軸を交差する割合が高いことが考えられる．MFCC や発話埋め込みに関する特徴量は，Praise シーンの方が大きい値を取得するものと，Non-praise シーンの方が大きい値を取得するものが存在した．今後，MFCC や発話埋め込みに関する特徴量が，どのような条件で Praise シーンもしくは Non-praise シーンで大きい値を取得するのか分析する必要がある．

4.4　褒め方の上手さを推定する機械学習モデルの構築

本節では，褒め手および受け手の言語・非言語行動から褒め方の上手さを推定する機械学習モデルについて述べる．具体的には，視覚的，韻律的，言語的特徴量を用いて褒め方の上手さを推定する機械学習モデルを構築する取り組みを行う．視覚的，韻律的，言語的特徴量の内容や抽出方法は，4.3.1 項と同様である．

††*は Praise シーンと Non-praise シーンの間に 5%水準の有意差が認められた特徴量．

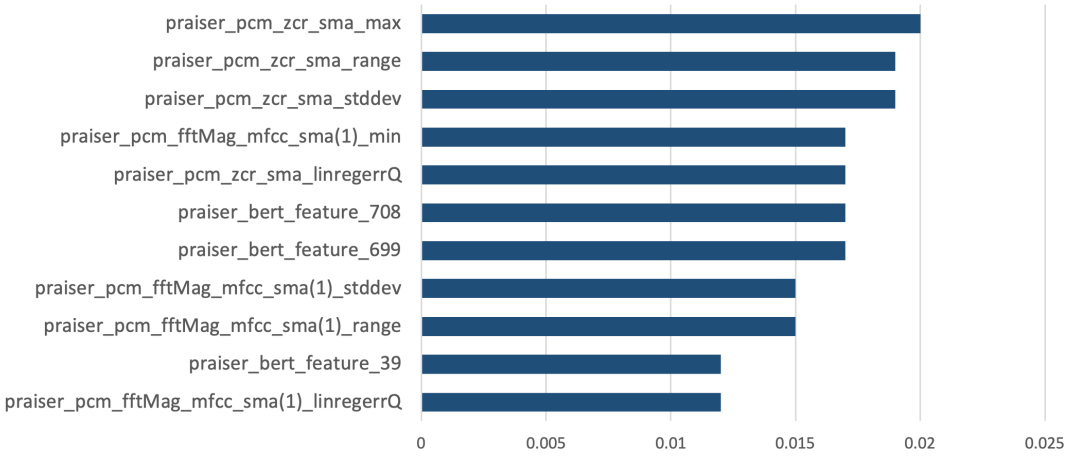

図 4.5: 褒める行為の検出における特徴量重要度（Gain）上位 10 件

4.4.1　機械学習モデルの構築

本項では，言語・非言語行動に関する特徴量から褒め方の上手さを推定する機械学習モデルについて述べる．本研究は，目的変数を 4.2.2.3 目で定義した Praise スコア低群，中群，高群とし，説明変数を 4.3.1 項で抽出した特徴量とする機械学習モデルを構築した．4.3.2 項と同様に Random Forests [118] を用いて機械学習モデルを構築し，Hyperopt [119] を用いてハイパーパラメータをチューニングした．さらに，モデルの性能の向上が止まるまで，特徴量重要度（Gain）の低い特徴量を削除する方法で特徴量を選択した．本項では，下記のタスクを 100 回繰り返して行った．

(1) データセットを，訓練データ 90%，テストデータ 10%に無作為に分ける．このとき，訓練データとテストデータにおいて各クラスに属するシーン数の割合が同じになるように分割した．

(2) 訓練データで構築したモデルを用いて，テストデータが属するクラスを分類する．

褒め方の上手さを推定するタスクにおいて，ベースラインモデルは，4.2.2.3 目で定義した Praise スコア低群，中群，高群の各シーン数の割合に合わせて，36%，28%，36%の確率で Praise スコア低群，中群，高群を推定するモデルを採用した．

4.4.2　機械学習モデルの性能

機械学習モデルの性能を表 4.7 に示す．本研究における検出性能を F 値で表す．表 4.7 の F 値は，各モデルを 100 回構築した際の重み付き F 値の平均値である．表 4.7 に示すように，ベースラインの検出性能（F 値）は 0.335 であった．本稿で提案した機械学習モデルの中で，最も検出性能が高かったモデルは M46 であり，検出性能（F 値）は 0.611 であった．M46 で得られたパラメータは次のとおりであった．

表 4.6: 褒める行為の検出における特徴量重要度上位 10 件の分布

特徴量	Praise シーン		Non-praise シーン
praiser_pcm_zcr_sma_max*	0.369 ± 0.063	$>$	0.224 ± 0.137
praiser_pcm_zcr_sma_range*	0.353 ± 0.067	$>$	0.205 ± 0.139
praiser_pcm_zcr_sma_stddev*	0.097 ± 0.025	$>$	0.053 ± 0.037
praiser_pcm_fftMag_mfcc_sma(1)_min*	-29.537 ± 3.978	$<$	-20.834 ± 9.632
praiser_pcm_zcr_sma_linregerrQ*	0.009 ± 0.004	$>$	0.004 ± 0.004
praiser_bert_feature_708*	-0.604 ± 0.225	$<$	-0.281 ± 0.293
praiser_bert_feature_699*	-0.259 ± 0.164	$<$	0.019 ± 0.296
praiser_pcm_fftMag_mfcc_sma(1)_stddev*	7.912 ± 1.807	$>$	4.935 ± 2.663
praiser_pcm_fftMag_mfcc_sma(1)_range*	30.626 ± 5.813	$>$	19.617 ± 11.002
praiser_bert_feature_39*	0.461 ± 0.247	$>$	0.108 ± 0.349
praiser_pcm_fftMag_mfcc_sma(1)_linregerrQ*	56.453 ± 25.594	$>$	26.317 ± 25.302

木の本数（n_estimators）： 65

木の深さ（max_depth）： 5

4.4.3　褒め方の上手さの推定に寄与するモダリティについて

本項では，褒め方の上手さの推定に寄与したモダリティについて述べる．

4.4.3.1　ユニモーダル特徴量を用いた推定

ユニモーダル特徴量を用いて褒め方の上手さを推定した場合（M1-M6），ベースラインモデルよりも推定性能を向上させることができた．このことから，本研究で利用した褒め手のすべての特徴量と受け手のすべての特徴量が推定に有用であると考えられる．その中で，最も高い推定性能を達成したのは，褒め手の韻律的特徴量を用いたモデル（M2）であり，推定性能（F 値）は $F = 0.512$ であった．上記より，褒め手の韻律的特徴量が褒め方の上手さを推定する上で重要な特徴量であると考えられる．

4.4.3.2　褒め手の特徴量を用いた推定

褒め手の特徴量を用いて褒め方の上手さを推定した場合（M1-M4, M7-M17），ベースラインモデルよりも推定性能を向上させることができた．その中で，最も高い推定性能を達成したのは，褒め手の韻律的特徴量と発話埋め込み特徴量を用いたモデル（M11）であり，推定性能（F 値）は $F = 0.579$ であった．このモデルは，褒め手のすべての特徴量を

表 4.7: 褒め方の上手さを推定する機械学習モデルの性能

	褒め手				受け手				褒め手				受け手		
	非言語		言語		非言語		F 値		非言語		言語		非言語		F 値
	視覚的	韻律的	言語的（品詞）	言語的（BERT）	視覚的	韻律的			視覚的	韻律的	言語的（品詞）	言語的（BERT）	視覚的	韻律的	
Baseline							0.335	M32	✓			✓		✓	0.517
M1	✓						0.453	M33	✓				✓	✓	0.427
M2		✓					0.512	M34		✓	✓		✓		0.501
M3			✓				0.392	M35		✓	✓			✓	0.493
M4				✓			0.526	M36		✓		✓	✓		0.566
M5					✓		0.393	M37		✓		✓		✓	0.593
M6						✓	0.430	M38		✓			✓	✓	0.505
M7	✓	✓					0.533	M39			✓	✓	✓		0.528
M8	✓		✓				0.472	M40			✓	✓		✓	0.557
M9	✓			✓			0.564	M41			✓		✓	✓	0.366
M10		✓	✓				0.494	M42				✓	✓	✓	0.540
M11		✓		✓			0.579	M43	✓	✓	✓		✓		0.521
M12			✓	✓			0.530	M44	✓	✓	✓			✓	0.532
M13	✓	✓	✓				0.519	M45	✓	✓		✓	✓		0.540
M14	✓	✓		✓			0.540	M46	✓	✓		✓		✓	**0.611**
M15	✓		✓	✓			0.542	M47	✓	✓			✓	✓	0.535
M16		✓	✓	✓			0.544	M48	✓		✓	✓	✓		0.529
M17	✓	✓	✓	✓			0.537	M49	✓		✓	✓		✓	0.531
M18					✓	✓	0.458	M50	✓		✓		✓	✓	0.418
M19	✓				✓		0.474	M51	✓			✓	✓	✓	0.509
M20	✓					✓	0.401	M52		✓	✓	✓	✓		0.531
M21		✓			✓		0.502	M53		✓	✓	✓		✓	0.509
M22		✓				✓	0.536	M54		✓	✓		✓	✓	0.556
M23			✓		✓		0.437	M55		✓		✓	✓	✓	0.547
M24			✓			✓	0.338	M56			✓	✓	✓	✓	0.510
M25				✓	✓		0.539	M57	✓	✓	✓	✓	✓		0.552
M26				✓		✓	0.537	M58	✓	✓	✓	✓		✓	0.549
M27	✓	✓			✓		0.580	M59	✓	✓	✓		✓	✓	0.553
M28	✓	✓				✓	0.526	M60	✓	✓		✓	✓	✓	0.579
M29	✓		✓		✓		0.461	M61	✓		✓	✓	✓	✓	0.587
M30	✓		✓			✓	0.303	M62		✓	✓	✓	✓	✓	0.516
M31	✓			✓	✓		0.541	M63	✓	✓	✓	✓	✓	✓	0.531

用いたモデル（M17）よりも推定性能が高かったため，褒め手のすべてのモダリティを用いなくても褒め方の上手さを十分に推定できることが示唆された．

4.4.3.3　受け手の特徴量を用いた推定

受け手の特徴量を用いて褒め方の上手さを推定した場合（M5，M6，M18），ベースラインよりも推定性能を向上させることができた．その中で，最も高い推定性能を達成したのは，受け手の視覚的特徴量と韻律的特徴量を用いたモデル（M18）であり，推定性能（F 値）は $F = 0.458$ であった．このモデルは，ベースラインよりも推定性能が高かったものの，褒め手の特徴量を用いたモデルよりも推定性能が高くないため，受け手のみでは褒め方の上手さを推定することは難しいと考えられる．

4.4.3.4　褒め手と受け手の特徴量を用いた推定

両者（褒め手と受け手）の特徴量を用いて褒め方の上手さを推定した場合（M19-M63），褒め手の視覚的特徴量と品詞特徴量と受け手の韻律的特徴量を用いたモデル（M30）を除いて，ベースラインモデルよりも推定性能を向上させることができた．その中で，最

も高い推定性能を達成したモデルは，褒め手の視覚的特徴量，韻律的特徴量，発話埋め込み特徴量と受け手の韻律的特徴量を用いたモデル（M46）であり，推定性能（F 値）が $F = 0.611$ であった．このことから，褒め手と受け手の両者の特徴量を用いることが褒め方の上手さの推定に有用であると考えられる．その中で，褒め手の視覚的特徴量，韻律的特徴量，発話埋め込み特徴量と受け手の韻律的特徴量に関するモダリティが，対話中の褒め方の上手さを推定する上で重要な振舞いとなることが考えられる．

全てのモデルの中で，推定性能（F 値）が最も高かったモデルは，褒め手の視覚的特徴量，韻律的特徴量，発話埋め込み特徴量と受け手の韻律的特徴量を用いたモデル（M46）であった．一方で，4.3.3 項の結果より，褒める行為を検出する際に最も推定性能が高かったモデルは褒め手の韻律的特徴量と発話埋め込み特徴量と受け手の視覚的特徴量と韻律的特徴量を用いていた．このことから，褒める行為を検出する際と褒め方の上手さを推定する際で重要なモダリティの組み合わせが異なることがわかった．褒め方の上手さを推定する際は，褒め手の視覚的特徴量が有用であったが，褒める行為の検出では，必ずしも褒め手の視覚的特徴量を用いる必要はないことが考えられる．上記より，人々がコミュニケーション中に褒める行為を行うために意識すべき振舞いと，褒め方の上手さを向上させるために意識すべき振舞いが異なることが考えられる．

4.4.3.5　褒め方の上手さの推定における特徴量重要度について

褒め方の上手さの推定に寄与したモダリティを明らかにするために，特徴量重要度（Gain）を算出した．最も推定性能の高かったモデル（M46）における特徴量重要度を図 4.6 に示す．図 4.6 より，褒め方の上手さの推定に寄与していた特徴量上位 10 件は，褒め手のゼロ交差率，メル周波数ケプストラム係数（MFCC），発話埋め込みに関する特徴量であった．これらの特徴量の分布を確認するために，Praise スコア低群と高群における各特徴量の平均値と標準偏差を表 4.8 に示す．さらに，各特徴量と Prise スコアの相関係数を表 4.8[‡‡]に示す．Praise スコア低群と高群の間で大小関係があるのか確認するために，対応のない t 検定を行った．検定の結果，褒め方の上手さの推定に寄与していた重要度上位 10 件の特徴量は，Praise スコア低群と高群の間に 5%水準の有意差が認められた．このことから，重要度上位 10 件の特徴量は，Praise スコア低群と高群の間に大小関係があると考えられる．表 4.8 より，重要度上位 10 件の特徴量は，Praise スコアと正の弱い相関を持つ特徴量もしくは負の弱い相関を持つ特徴量であることが確認できる．最も重要度の高かった特徴量は，ゼロ交差率の分散を表しており，Praise スコアと正の弱い相関がある．このことから，褒め手が上手く褒めることができているシーンにおいて，褒める発話を行う際の音声は音圧 0 である軸を交差する割合のばらつきが大きいことが考えられる．今後，重要度の高かったゼロ交差率，MFCC，発話埋め込みに関する特徴量が，どのような条件で Praise 低群もしくは高群で大きい値を取得するのか分析する必要がある．

[‡‡]*は Praise スコア低群と高群の間に 5%水準の有意差が認められた特徴量．

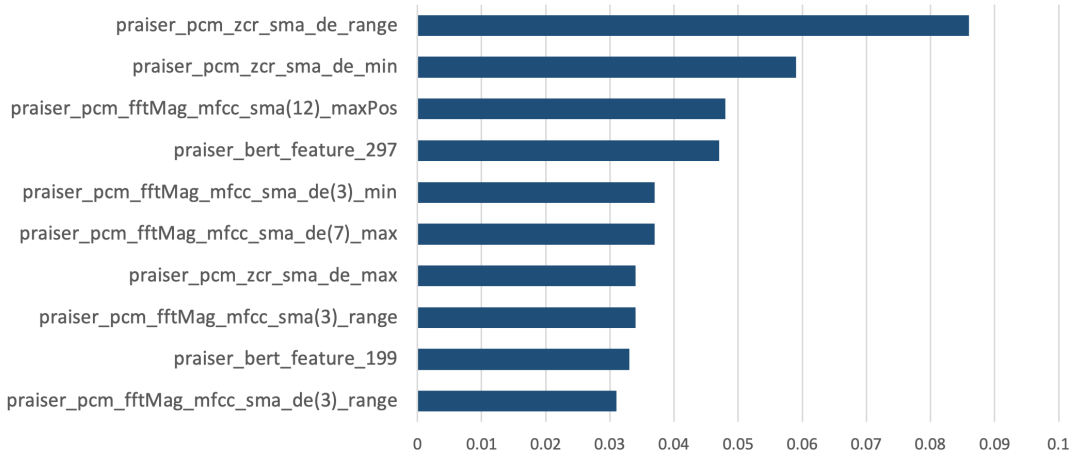

図 4.6: 褒め方の上手さの推定における特徴量重要度（Gain）上位 10 件

4.5　本章のまとめ

本章では，褒め手および受け手の言語・非言語行動から褒める行為の検出や褒め方の上手さの推定を行う機械学習モデルを構築した．

はじめに，実対話データと褒める行為に関する評価値が記録された対話コーパスを構築した．この対話コーパスは，実際の 2 者対話を 17 組分，計 255 分間記録している．加えて，5 名の評価者が褒め手が対話相手のことを褒めているか，褒め手の褒め方は上手いと感じたかについて評価した結果を記録している．

次に，褒める行為を検出する機械学習モデルと褒め方の上手さを推定する機械学習モデルを構築した．機械学習モデルで用いる特徴量は，褒め手と受け手の言語・非言語行動から視覚的・韻律的・言語的特徴量を抽出した．4.3.3 項より，褒める行為を検出する機械学習モデルを構築した結果，褒め手の韻律的特徴量と発話埋め込み特徴量と受け手の視覚的特徴量と韻律的特徴量を用いたモデル（M55，$F = 0.827$）が最も検出性能を向上させることができた．このことから，褒め手の韻律的特徴量と発話埋め込み特徴量と受け手の視覚的特徴量と韻律的特徴量に関するモダリティが，対話中の褒める行為を検出する上で重要な振舞いであることが示唆された．4.4.2 項より，褒め方の上手さを推定する機械学習モデルを構築した結果，褒め手の視覚的特徴量，韻律的特徴量，発話埋め込み特徴量と受け手の韻律的特徴量を用いたモデル（M46，$F = 0.611$）が最も推定性能を向上させることができた．このことから，褒め手の視覚的特徴量，韻律的特徴量，発話埋め込み特徴量と受け手の韻律的特徴量に関するモダリティが，対話中の褒め方の上手さを推定する上で重要な振舞いであることが示唆された．

本取り組みでは上記の結果が得られたが，複数の制約が存在する．はじめに，本取り組みでは 2 者対話を収録する際，褒め手には対話相手を積極的に褒めるように指示をした．しかし，過半数のアノテータが褒めているシーンであると判断したシーン（Praise シーン）の件数は，228 件と少なかった．今後，機械学習モデルの検出や推定性能を向上させ

表 4.8: 褒め方の上手さの推定における特徴量重要度上位 10 件の分布と Praise スコアとの相関係数

特徴量	スコア低群		スコア高群	相関係数
praiser_pcm_zcr_sma_de_range*	0.108 ± 0.032	$<$	0.146 ± 0.028	0.460
praiser_pcm_zcr_sma_de_min*	-0.054 ± 0.017	$>$	-0.074 ± 0.014	-0.451
praiser_pcm_fftMag_mfcc_sma(12)_maxPos*	59.134 ± 56.768	$<$	189.765 ± 157.624	0.455
praiser_bert_feature_297*	-0.329 ± 0.146	$<$	-0.124 ± 0.168	0.436
praiser_pcm_fftMag_mfcc_sma_de(3)_min*	-2.989 ± 1.155	$>$	-4.388 ± 1.062	-0.443
praiser_pcm_fftMag_mfcc_sma_de(7)_max*	3.165 ± 0.966	$<$	4.435 ± 1.092	0.501
praiser_pcm_zcr_sma_de_max*	0.053 ± 0.019	$<$	0.073 ± 0.017	0.412
praiser_pcm_fftMag_mfcc_sma(3)_range*	24.025 ± 7.678	$<$	32.296 ± 6.231	0.411
praiser_bert_feature_199*	0.206 ± 0.183	$>$	0.008 ± 0.149	-0.376
praiser_pcm_fftMag_mfcc_sma_de(3)_range*	6.318 ± 2.080	$<$	8.848 ± 1.866	0.452

るためには，新たに対話データを収録し，Praise シーン数を増やすことを検討する必要がある．

次に，本取り組みでは，褒め手が褒めているタイミング（Praise シーン）から受け手の特徴量を抽出した．そのため，受け手の特徴量を用いたモデルは，ベースラインよりも推定性能が高かったものの，褒め手の特徴量を用いたモデルよりも推定性能が高くなることはなかった．既存研究において，受け手は褒められることでポジティブな表出を多くすることが報告されている [69]．褒め手が褒めた後（Praise シーンの後）のタイミングなどから受け手の特徴量を抽出することで，受け手の特徴量を用いたモデルの推定性能が向上するのではないかと考えられる．今後，機械学習モデルを改善する際は，受け手の特徴量の抽出範囲を再検討する必要がある．

第5章　聞き手の理解度を自動推定する取り組み

5.1　聞き手の理解度を自動推定する重要性

聞き手の理解を表す行為は，話し手の発話に対して理解を示し，コミュニケーションを促進させるために重要であると考えられている．2.2.2 項で述べたように，聞き手のフィードバックは，聞き手が話し手の話を聞いていることを表す合図や，話し手の発話内容を理解している，もしくは理解しようとしていることを表す合図と言われており [8,9]，コミュニケーション中に聞き手は理解を表すことは必要不可欠である．しかし，コミュニケーションが苦手な人やコミュニケーションスキルが乏しい人は，自身が聞き手であった場合に理解できているもしくは理解できていないことを対話相手（話し手）に伝えられていない可能性がある．さらに自身が話し手であった場合に自身の発言に対して対話相手（聞き手）が理解をしている，もしくは理解をしていないことを把握できていない可能性がある．コミュニケーションを行っている最中に，コミュニケーションが苦手な人やコミュニケーションスキルが乏しい人が自身が理解していることを相手に伝えられているか把握することや，自身の発言を相手が理解をしているか把握することは難しいと考えられる．

この問題を解決するためには，コミュニケーションを行っている最中に自身が理解していることを相手に伝えられているか，自身の発言を相手が理解をしているか把握できるようにする必要がある．これを実現するためには，話し手と聞き手以外の参加者が聞き手が理解できているか否かをコミュニケーション中に伝える方法が考えられる．しかし，話し手と聞き手以外の参加者が常に近くにいるとは限らない．さらには，1 対 1 の状況でコミュニケーションを行いたい場合も存在する．

上記をふまえ，コミュニケーション中に聞き手が理解できているか否かを，自動的に判定する方法が有効であると考えられる．自動的に判定することで，話し手と聞き手以外の参加者がいない場合や 1 対 1 の状況でコミュニケーションを行いたい場合でも，自身が理解していることを相手に伝えられているか，自身の発言を相手が理解をしているか把握できるようになると考えられる．本研究は，聞き手が理解できているか否かを，自動的に判定する初期的な取り組みとして，聞き手の言語・非言語行動から聞き手の理解度を推定する取り組みを行う [54].

5.2 節では，聞き手の理解度を推定するための対話コーパスについて説明する．5.3 節では，聞き手の理解度を推定する機械学習モデルについて説明する．5.4 節では，聞き手の理解度を推定する取り組みについて総括する．

5.2　聞き手の理解度を推定するための対話コーパス

本節では，聞き手の理解度を推定するための対話コーパスについて述べる．本研究は，既存研究で作成された対話コーパスを利用した [121]．この対話コーパスには，2 者対話における話者の言語・非言語行動と聞き手の理解を表す行為の評価値が記録されている．

5.2.1　2 者対話データ

本項では，2 者対話データについて述べる．本研究は，既存研究 [121] で記録した 2 者対話データを利用した．この対話データは，対話参加者の正面に設置したビデオカメラで撮影した映像データ，対話参加者に取り付けたマイクで録音した音声データ，音声データから書き起こされた発話内容を記録している．2 者対話の参加者は，合計で 26 名（異なるペアが 13 組）であり，初対面の 20 代から 50 代日本人男女である．聞き手が対話相手の話に対して理解していること表現しているデータをより多く収集するため，ストーリーテリングをタスクとしている．具体的には，一方の参加者（話し手）が他方の参加者（聞き手）にアニメ「トムとジェリー」の内容を説明するタスク [121] を行っている．発話の単位は，Inter-pausal units (IPU) [99] を用いており，沈黙時間が 200ms 未満の連続した音声区間を 1 つとしている．この対話データでは，合計 7,805 件（話し手：4,940 件，聞き手：2,865 件）の発話（IPU）が記録されている．

5.2.2　アノテーション

本項では，理解度の定義と理解度のアノテーションについて述べる．理解度のアノテーションを始めるにあたり，理解度を定義する．本研究は，聞き手が対話相手の話す内容を把握し，理解している様子に着目する．本研究は，理解度を聞き手が話し手の発話を理解している様子の度合いと定義する．

対話参加者の内部情報を第三者が評価する方法が，多くの類似研究で用いられている [103–107]．さらに，第三者による客観的な観察に基づく評価と参加者本人による主観的な評価には，一定の相関関係があることも報告されている [102]．これらのことから，本研究では第三者が聞き手の理解度を評価する方法を採用した．

映像や音声データに対して注釈付けを行うツールである ELAN [98] を用いて，対話に参加していない第三者のアノテータ 3 名が，聞き手の理解度の評価を行った．アノテータは，話し手と聞き手の映像データと音声データを視聴して，任意の評価区間を設定して次の評価を行った．理解度（相手の話す内容をどれくらい把握し，理解できているか）を 5 段階のリッカート尺度（-2：全く理解していない，-1：あまり理解していない，0；中立の状態，+1；少し理解している，+2；とても理解している）で評価を行った．このとき，対話開始から終了まで，評価のない時間帯（区間）がないようにすることを指示した．アノテーションされた対話の一部を図 5.1 に示す．本研究では，聞き手の発話終了時点でアノテータが評価したスコアを聞き手の理解度として用いた．理解度の分布を図 5.2 に示す．

アノテータ間の評価の一致度を確認するために，級内相関係数（Intraclass correlation coefficients, ICC）[109] を利用して，アノテータの評価の一致率を算出した．聞き手の発話区間における 3 名のアノテータが評価した理解度の一致率を算出したところ，$ICC(2,k) = 0.546$ であり，アノテータ間の評価の一致度は中程度 [110] であることがわかった．本研究と同じように第三者が評価している研究 [122] の評価の一致率と同程度であり，理解度の評価結果は信頼性の高いデータであることが示唆された．

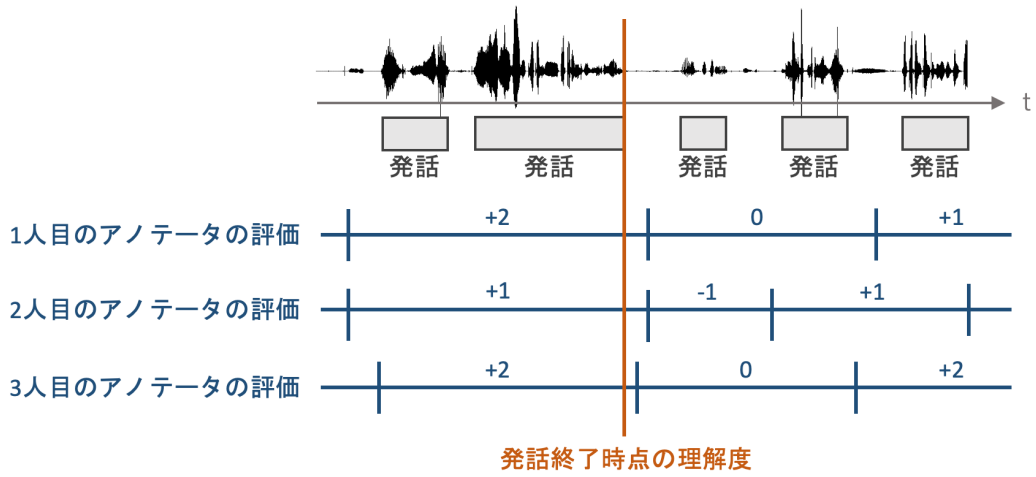

図 5.1: 理解度のアノテーション方法の例

上記のアノテーション結果に基づいて，理解度を 3 つのクラスに分割する．

理解クラス： 理解している（+1 もしくは+2）と評価したアノテータが 2 名以上の発話（計 1,745 シーン）

非理解クラス： 理解していない（-1 もしくは-2）と評価したアノテータが 2 名以上の発話（計 96 シーン）

中立クラス： 上記の理解，非理解クラスに該当しない発話（計 1,020 シーン）

5.3　聞き手の理解度を推定する機械学習モデルの構築

本節では，聞き手の言語・非言語行動から聞き手の理解度を機械学習を用いて推定する取り組みを行う．具体的には，実対話と聞き手の理解度を記録した対話コーパスを利用し，聞き手の視覚的，韻律的，言語的特徴量から聞き手の理解度を推定する機械学習モデルを構築する取り組みを行う．

5.3.1　特徴量抽出

本項では，5.2.1 項で記録した対話データから聞き手の視覚的，韻律的，言語的特徴量を抽出した．特徴量は，聞き手の発話区間から抽出した．

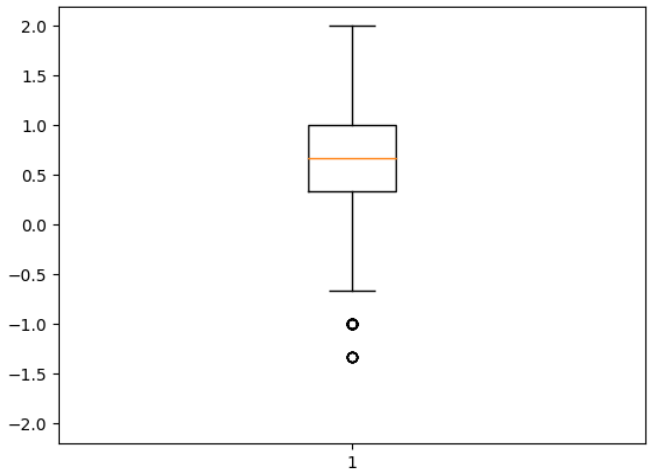

図 5.2: 理解度の分布（N=2,861）

5.3.1.1　視覚的特徴量

OpenFace [112] を用いて，対話参加者（聞き手）の正面に設置したビデオカメラで撮影した映像データから頭部，視線，表情（Action Units [84]）に関する特徴量を抽出した．抽出した特徴量は，合計 88 次元である．特徴量の詳細は次のとおりである．

頭部： 聞き手の発話区間における頭部の x 軸，y 軸，z 軸周りの回転角度の分散，中央値，10 パーセンタイル値，90 パーセンタイル値を用いた．

視線： 聞き手の発話区間における視線の x 軸，y 軸方向の角度の分散，中央値，10 パーセンタイル値，90 パーセンタイル値を用いた．

Action Units： 聞き手の発話区間における各 Action Units（表 5.1）の強度の分散，中央値，10 パーセンタイル値，90 パーセンタイル値を用いた．

5.3.1.2　韻律的特徴量

openSMILE [114] を用いて，対話参加者（聞き手）に取り付けたマイクで録音した音声データから韻律の代表的な特徴量と発話時間に関する特徴量を抽出した．表 5.2 に示した韻律の代表的な特徴量につき，表 5.3 に示す統計量を算出した特徴量を抽出した．これらの特徴量は，openSMILE の Python ライブラリで提供されたものの中から，4.3.1.2 目で用いられている特徴量のみにした．韻律的特徴量は，合計 192 次元であった．

表 5.1: OpenFace での Action Units の内容

項目	内容	項目	内容
AU01	眉の内側を上げる	AU14	笑窪を作る
AU02	眉の外側を上げる	AU15	唇の両端を下げる
AU04	眉を下げる	AU17	顎を上げる
AU05	上瞼を上げる	AU20	唇の両端を横に引く
AU06	頬を持ち上げる	AU23	唇を固く閉じる
AU07	瞳を緊張させる	AU25	顎を下げずに唇を開く
AU09	鼻に皺を寄せる	AU26	顎を下げて唇を開く
AU10	上唇を上げる	AU45	瞬きをする
AU12	唇の両端を引き上げる		

表 5.2: 韻律の代表的な特徴量

特徴量	内容
pcm_intensity_sma	正規化された強度の値
mfcc_sma[1]–[12]	1〜12 次のメル周波数ケプストラム係数
pcm_zcr_sma	ゼロ交差率
voiceProb_sma	声である確率
F0_sma	基本周波数

5.3.1.3　言語的特徴量

MeCab [116] を用いて，5.2.1 項の発話内容から，品詞の出現頻度と単語数を算出し，特徴量として用いた．抽出した品詞を表 5.4 に示す．言語的特徴量は，全部で 37 次元であった．

5.3.2　機械学習モデルの構築

5.3.1 項で抽出した聞き手の視覚的，韻律的，言語的特徴量から，聞き手の理解度を推定するモデルを構築した．目的変数を 5.2.2 項で定義した理解，非理解，中立クラスとし，説明変数を 5.3.1 項で抽出した特徴量とする機械学習モデルを構築した．Random Forests [118] を用いて機械学習モデルを構築し，Hyperopt [119] を用いて，ハイパーパラメータをチューニングした．各パラメータの探索範囲は次のとおりである．

木の本数（n_estimators）： 10〜100

表 5.3: 抽出に用いた統計量

統計量	内容	統計量	内容
_max	最大値	_stddev	標準偏差
_min	最小値	_linregc1	線形近似の勾配
_range	最大値と最小値の差	_linregc2	線形近似のオフセット
_maxPos	最大値の絶対位置	_linregerrQ	線形近似の二乗誤差
_minPos	最小値の絶対値位置	_skewness	歪度
_amean	平均値	_kurtosis	尖度

木の深さ（max_depth）： 3〜10

ハイパーパラメータのチューニングは，訓練データ 90%とテストデータ 10%に分割されたデータを用いて，1,000 回行った．1,000 回の試行結果に基づいて，最も性能の良いパラメータを用いることとした．さらに，モデルの性能の向上が止まるまで，特徴量重要度（Gain）の低い特徴量を削除する方法で特徴量を選択した．評価指標は，重み付きの適合率，再現率，F 値を用いた．本研究では，次のタスクを 100 回繰り返して行った．

(1) データセットを，訓練データ 90%，テストデータ 10%に無作為に分ける．このとき，訓練データとテストデータにおいて各クラスに属するシーン数の割合が同じになるように分割した．

(2) 訓練データで構築したモデルを用いて，テストデータが属するクラスを分類する．

ベースラインモデルは，5.2.2 項で定義した理解，非理解，中立クラスに属するシーン数の割合に合わせて，61%，36%，3%の確率で理解，非理解，中立クラスを推定するモデルを採用した．

5.3.3　機械学習モデルの性能

機械学習モデルの性能を表 5.5 に示す．表 5.5 の各指標は，各モデルを 100 回繰り返し構築した際の平均値である．表 5.5 に示すように，ベースラインの推定性能（F 値）は 0.506 であった．本稿で提案した機械学習モデルの中で，最も検出性能が高かったモデルは M7 であり，推定性能（F 値）は 0.599 であった．M7 で得られたパラメータは次のとおりであった．

木の本数（n_estimators）： 24

木の深さ（max_depth）： 3

表 5.4: 抽出した品詞

大分類	小分類
フィラー	フィラー
感動詞	感動詞
形容詞	形容詞
助詞	助詞（格助詞），助詞（副助詞），助詞（連体化）， 助詞（接続助詞），助詞（特殊），助詞（副詞化）， 助詞（副助詞），助詞（副助詞／並立助詞／終助詞）， 助詞（並立助詞），助詞（連体化），
助動詞	助動詞
接続詞	接続詞
動詞	動詞（自立），動詞（接尾），動詞（非自立）
名詞	名詞（サ変接続），名詞（ナイ形容詞語幹），名詞（一般）， 名詞（引用文字列），名詞（形容動詞語幹），名詞（固有名詞）， 名詞（数），名詞（接続詞的），名詞（接尾），名詞（代名詞）， 名詞（動詞非自立的），名詞（特殊），名詞（非自立）， 名詞（副詞可能）
連体詞	連体詞
副詞	副詞
接頭詞	接頭詞
記号	記号

5.3.4　理解度の推定に寄与するモダリティについて

本項では，聞き手の理解度の推定に寄与したモダリティについて述べる．

5.3.4.1　ユニモーダル特徴量を用いた予測

5.3.3 項より，ユニモーダル特徴量を用いて聞き手の理解度を推定した場合（M1〜M3），視覚的特徴量を用いたモデル（M1）と韻律的特徴量を用いたモデル（M2）はベースラインモデルよりも性能を向上させることができた．このことから，聞き手の視覚的，韻律的特徴量は聞き手の理解度を推定に有用であると考えられる．その中で，最も高い推定性能を達成したのは視覚的特徴量を用いたモデル（M1）であり，推定性能（F 値）は $F = 0.580$ であった．上記より，聞き手の視覚的特徴量に関するモダリティが，聞き手の理解度を推定する上で重要なモダリティであると考えられる．

表 5.5: 理解度を推定する機械学習モデルの性能（N=100）

	視覚的	韻律的	言語的	適合率	再現率	F 値
Baseline				0.507	0.506	0.506
M1	✓			0.631	0.548	0.580
M2		✓		0.625	0.539	0.572
M3			✓	0.619	0.446	0.486
M4	✓	✓		0.633	0.567	0.592
M5	✓		✓	0.616	0.531	0.564
M6		✓	✓	0.622	0.535	0.568
M7	✓	✓	✓	**0.640**	**0.575**	**0.599**

5.3.4.2　マルチモーダル特徴量を用いた予測

マルチモーダル特徴量を用いて聞き手の理解度を推定した場合（M4〜M7），ベースラインモデルよりも推定性能を向上させることができた．その中で，最も高い推定性能を達成したのは視覚的，韻律的，言語的特徴量を用いたモデル（M7）であり，推定性能（F 値）は $F = 0.599$ であった．このことから，聞き手の視覚的，韻律的，言語的特徴量を組み合わせることが聞き手の理解度の推定性能の向上に有用にあると考えられる．

5.3.4.3　理解度の推定における特徴量重要度について

推定に寄与したモダリティを明らかにするために，5.3.2 項で構築したモデルにおける特徴量重要度（Gain）を算出した．最も推定性能の高かったモデル（M7）における特徴量の重要度を図 5.3 に示す．聞き手の理解度に寄与していた特徴量上位 10 件は，MFCC，AU04，AU12 に関する特徴量であることが確認できる．AU04 は，頬を持ち上げる動きを表す特徴量である．AU12 は，唇の両端を引き上げる動きを表す特徴量である．このことから，M57 で用いた特徴量（聞き手の視覚的，韻律的，言語的特徴量）の中で，MFCC，AU04，AU12 に関する特徴量が理解度の推定に寄与していたことが確認された．

5.4　本章のまとめ

本章では，聞き手の言語・非言語行動から聞き手の理解度を推定する機械学習モデルを構築する取り組みを行った．本取り組みでは，実対話データと聞き手の理解している様子に関する評価値が記録された対話コーパスを用いた．この対話コーパスには，実際の 2 者対話が 13 組分記録されている．加えて，3 名の評価者が聞き手が相手の話す内容をどれくらい把握し，理解できているかについて評価した結果が記録されている．

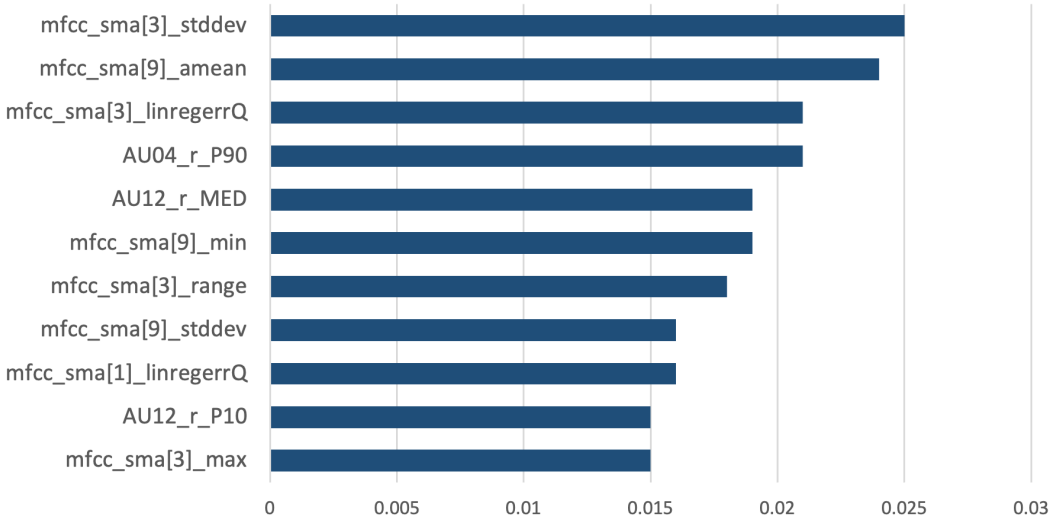

図 5.3: 理解度の推定における特徴量重要度（Gain）上位 10 件

聞き手の言語・非言語行動から聞き手の理解度を推定する機械学習モデルを構築した．機械学習モデルで用いた特徴量は，聞き手の視覚的，韻律的，言語的特徴量を用いた．5.3.3 項より，聞き手の理解度を推定する機械学習モデルを構築した結果，聞き手の視覚的，韻律的，言語的特徴量を用いたモデル（M7，$F = 0.599$）が，最も推定性能を向上させることができた．このことから，聞き手の視覚的，韻律的，言語的特徴量に関するモダリティが，聞き手の理解度を推定する上で有用な振舞いであることが示唆された．

本取り組みでは上記の結果が得られたが，制約が存在する．本取り組みでは，聞き手の発話終了時点でアノテータが評価したスコアを聞き手の理解度として用いた．アノテーションを行う際，アノテータは任意の評価区間を設定して，聞き手の理解度を評価した．そのため，聞き手が発話をしていないときの理解度や聞き手が発話を開始した時点での理解度などは推定の対象外であった．コミュニケーション中に聞き手が理解できているか否かを自動的に判定するためには，聞き手が発話をしていないときや，聞き手が発話を開始したときの理解度も推定の対象とする必要があると考えられる．今後，聞き手の理解度を再定義した上で，聞き手の理解度を推定する必要がある．

第6章　聞き手の相槌機能を自動予測する取り組み

6.1　聞き手の相槌機能を自動予測する重要性

日常生活や社会活動でのコミュニケーションにおいて，聞き手が「うんうん」,「はい」,「すごい」などと言語的に相槌をすることや，頭部を動かしたり表情を変化させたりするなど非言語的に相槌をすることがある．このような相槌は，対話を成立させ，円滑に進めるために重要であると考えられている [7]．相槌は，継続を促す表現，理解を示す表現，同意をする表現，感情の表現，情報を要求する表現などの機能別に分類することができると考えられている [39]．聞き手が相槌機能をコミュニケーション中に適切に用いることで，コミュニケーションが円滑に進行することが期待される．

コミュニケーション中に発生する相槌に着目した研究の多くは，コミュニケーションの構造を分析し，聞き手が相槌を打つべきタイミングや聞き手の相槌が対話相手にどのような影響を与えるのか明らかにする取り組みを行っている [7, 9, 12, 13, 15–21]．これらの研究に基づいて，コミュニケーション中の話し手や聞き手の言語・非言語行動を用いて聞き手の相槌を推定する研究も多く行われている [70–80]．しかし，聞き手の相槌を推定する研究は，コミュニケーション中に聞き手の相槌が行われるか予測したり，どのタイミングで聞き手が相槌を打つべきなのか推定したりする取り組みが多く，聞き手が打つべき相槌機能を予測する取り組みは少ない．

そのため，コミュニケーションが苦手な人が自身が聞き手である際に，適切なタイミングで相槌を打つことができているのかを判断することはできるが，コミュニケーションの流れや内容に沿った相槌を打てているのか話し手の言語・非言語行動から判断することができないという問題がある．

この問題を解決するためには，話し手の言語・非言語行動から聞き手の相槌機能を自動的に予測できる必要がある．自動的に予測できるようになることで，コミュニケーションの流れや内容に適した相槌を提示できるようになると考えられる．そこで本研究では，話し手の言語・非言語行動から聞き手の相槌機能を予測する取り組みを行う [55]．

6.2 節では，聞き手の相槌機能を予測するための対話コーパスについて説明する．6.3 節では，聞き手の相槌機能を予測する機械学習モデルについて説明する．最後に，6.4 節では，聞き手の相槌機能を予測する取り組みについて総括する．

6.2　聞き手の相槌機能を自動予測するための対話コーパス

本節では，対話中の相槌を打つ行為を分析するための対話コーパスについて述べる．本研究は，既存研究で作成された対話コーパスを利用した [3, 121]．この対話コーパスには，2 者対話における話者の言語・非言語行動と聞き手の相槌機能ラベルが記録されている．

6.2.1　2 者対話データ

本項では，2 者対話データについて述べる．本研究は，既存研究 [121] で記録した 2 者対話データを利用した．この対話データは，対話参加者の正面に設置したビデオカメラで撮

影した映像データ，対話参加者に取り付けたマイクで録音した音声データ，音声データから書き起こされた発話内容を記録している．2 者対話の参加者は，合計で 26 名（異なるペアを 13 組）であり，初対面の日本人男女である．対話を記録する際，発話を含んだ相槌のデータをより多く収集するため，対話のタスクはストーリーテリングとしている．具体的には，一方の参加者（話し手）が他方の参加者（聞き手）にアニメ「トムとジェリー」の内容を説明するタスクを行っている．発話の単位は Inter-pausal units (IPU) [99] を用いており，沈黙時間が 200ms 未満の連続した音声区間を 1 つとしている．この対話データでは，合計 7,805 件（話し手：4,940 件，聞き手：2,865 件）の発話（IPU）が記録されている．

6.2.2　アノテーション

本研究は，既存研究 [3] で記録した相槌機能のラベルを利用した．6.2.2.1 目では，聞き手の相槌機能を表す相槌ラベルについて説明する．6.2.2.2 目では，相槌ラベルのアノテーション方法について説明する．

6.2.2.1　相槌ラベルについて

2.2.3 項で述べたように，聞き手のフィードバックは肯定的，否定的などの種類に分類することができる [6,12]．さらに，6.1 節で述べたように，聞き手の相槌は，継続，理解，同意，感情，要求などの機能に分類することができる [39]．これらの分類に基づき，Morikawa ら [3] は，表 6.1 に示す 9 種類の相槌機能を定義している．本研究は，Morikawa らが定義した相槌機能を相槌ラベル呼ぶ．

表 6.1: Morikawa ら [3] による相槌機能

ラベル	種類	説明
N	Neutral word	話し手への感情を含まない応答
P	Positive word	話し手への肯定的な応答
NP	Non-positive word	話し手への否定的または悩んでいるような応答
E	Emotional word	感情の動きを表しているような応答
A	Anticipation	話し手の話題を先取りしている応答
C	Confirmation	確認を促す，質問するような応答
R	Repetition of speaker’s utterance	話し手の発言を繰り返す応答
S	Summary of speaker’s utterance	話し手の発話の要約，および言い換えをしているような応答
O	Other	聞き手の感想や独り言など，他に該当するラベルがない応答

6.2.2.2　相槌ラベルのアノテーションについて

6.2.2.1 目で述べた相槌ラベルのアノテーション方法について説明する．対話に参加していない第三者のアノテータ 3 名が，対話参加者（話し手と聞き手）の正面に設置したビデオカメラで撮影した映像データと，対話参加者（話し手と聞き手）に取り付けたマイクで録音した音声データを視聴し，聞き手の発話ごとに相槌ラベルを記録している．相槌ラベルは，1 つの発話に対して複数記録されている場合もある．アノテーション方法の例を図 6.1 に示す．アノテータの判定の一致度を評価するために Fleiss のカッパ係数 [123] を用いて一致率を算出した．その結果，一致率は 0.712 であり，高い一致度であることが確認できた．

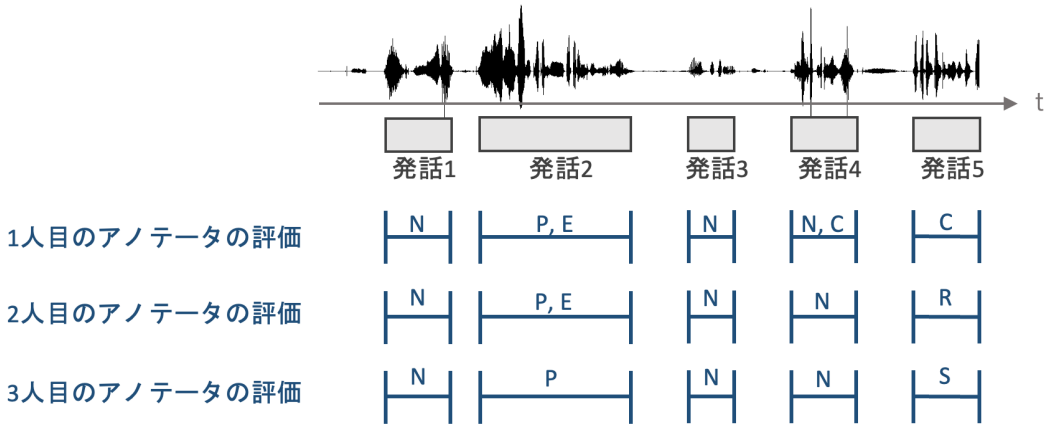

図 6.1: 相槌ラベルのアノテーションの例

6.3　聞き手の相槌機能を予測する機械学習モデルの構築

本節では，話し手の言語・非言語行動から相槌機能を予測する機械学習モデルについて述べる．具体的には，話し手の視覚的，韻律的，言語的特徴量を用いて聞き手の相槌機能を予測する機械学習モデルを構築する取り組みを行う．

6.3.1　特徴量抽出

本項では，6.2.1 節で記録した対話データから話し手の視覚的，韻律的，言語的特徴量を抽出した．具体的には，話し手の発話区間における話し手の視覚的，韻律的，言語的特徴量を抽出した（図 6.2）．

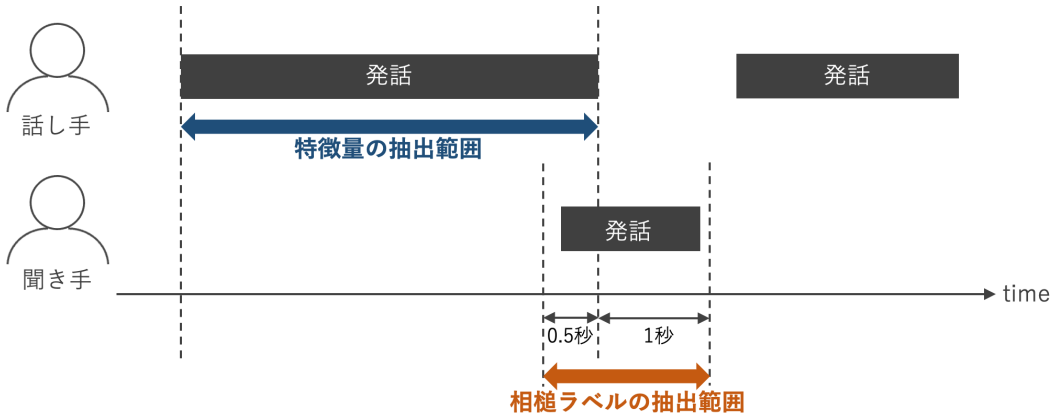

図 6.2: 相槌機能の予測における特徴量抽出範囲と相槌ラベル抽出範囲

6.3.1.1　視覚的特徴量

本研究は，OpenFace [112] を用いて，対話参加者（話し手）の正面に設置したビデオカメラで撮影した映像データから頭部，視線，表情（Action Units [84]）に関する特徴量を抽出した．抽出した特徴量は，合計 88 次元である．特徴量の詳細は次のとおりである．

頭部： 発話区間の前後 1 秒ずつを含む範囲における頭部の x 軸，y 軸，z 軸周りの回転角度の分散，中央値，10 パーセンタイル値，90 パーセンタイル値を用いた．

視線： 発話区間の前後 1 秒ずつを含む範囲における視線の x 軸，y 軸方向の角度の分散，中央値，10 パーセンタイル値，90 パーセンタイル値を用いた．

Action Units： 発話区間の前後 1 秒ずつを含む範囲における各 Action Units（表 6.2）の強度の分散，中央値，10 パーセンタイル値，90 パーセンタイル値を用いた．

6.3.1.2　韻律的特徴量

openSMILE [114] を用いて，対話参加者（話し手）に取り付けたマイクで録音した音声データから韻律の代表的な特徴量と発話時間に関する特徴量を抽出した．表 6.3 に示した韻律の代表的な特徴量につき，表 6.4 に示す統計量を算出した特徴量を抽出した．さらに，これらすべての特徴量の一次微分した特徴量（_de）を抽出した．これらの特徴量は，openSMILE の Python ライブラリで提供されたものである．韻律的特徴量は，合計 336 次元であった．

表 6.2: OpenFace での Action Units の内容

項目	内容	項目	内容
AU01	眉の内側を上げる	AU14	笑窪を作る
AU02	眉の外側を上げる	AU15	唇の両端を下げる
AU04	眉を下げる	AU17	顎を上げる
AU05	上瞼を上げる	AU20	唇の両端を横に引く
AU06	頬を持ち上げる	AU23	唇を固く閉じる
AU07	瞳を緊張させる	AU25	顎を下げずに唇を開く
AU09	鼻に皺を寄せる	AU26	顎を下げて唇を開く
AU10	上唇を上げる	AU45	瞬きをする
AU12	唇の両端を引き上げる		

表 6.3: 韻律の代表的な特徴量

特徴量	内容
pcm_intensity_sma	正規化された強度の値
pcm_loudness_sma	正規化された強度に 0.3 乗した値
mfcc_sma[1]–[12]	1〜12 次のメル周波数ケプストラム係数
lspFreq_sma[0]–[7]	8 つの LPC 係数から計算される周波数
pcm_zcr_sma	ゼロ交差率
voiceProb_sma	声である確率
F0_sma	基本周波数
F0env_sma	基本周波数のエンペローブ

6.3.1.3　言語的特徴量

MeCab [116] を用いて，6.2.1 項の発話内容から，品詞の出現頻度と単語数を算出し，特徴量として用いた．抽出した品詞を表 6.5 に示す．言語的特徴量は，全部で 37 次元であった．

6.3.2　機械学習モデルの構築

本研究では，話し手の発話終了前後における聞き手の相槌機能を予測する機械学習モデルを構築した．目的変数を 6.2.2.2 目で判定した相槌ラベルとし，説明変数を 6.3.1 項で抽出した特徴量とする機械学習モデルを構築した．

本研究は，相槌機能を予測する取り組みのファーストステップとして，話し手の発話終

表 6.4: 抽出に用いた統計量

統計量	内容	統計量	内容
_max	最大値	_linregc1	線形近似の勾配
_min	最小値	_linregc2	線形近似のオフセット
_range	最大値と最小値の差	_linregerrA	線形近似と実際の値の誤差の差
_maxPos	最大値の絶対位置	_linregerrQ	線形近似の二乗誤差
_minPos	最小値の絶対値位置	_skewness	歪度
_amean	平均値	_kurtosis	尖度
_stddev	標準偏差		

了前後に行われる聞き手の相槌に着目する．そのため，話し手の発話終了 0.5 秒前から 1 秒後の範囲を相槌ラベルの抽出範囲とし（図 6.2），この範囲内に存在する聞き手の発話区間に付与されている相槌ラベルを予測対象とした．さらに，本研究では，3 名のアノテータのうち 2 名以上が同じ判定をしている相槌ラベルのみを予測対象とした．図 6.1 の発話 5 のように，3 名とも別々の相槌ラベルを付与している場合は今回は予測の対象外とした．予測対象となる各相槌ラベルの数を表 6.6 に示す．

機械学習モデルの構築では，ニューラルネットワークモデル（付録 A.1）を用いた．本研究は，本研究が属する分野においてニューラルネットワークモデルを用いている研究事例 [38, 74] を参考にした．本研究で構築したモデルは，視覚的，韻律的，言語的特徴量を入力とし，9 種類の相槌ラベルを予測するモデルである．モデルの処理の流れを図 6.3 に示す．T はデータセットのサンプルサイズ，FC は全結合層を表す．まず，各特徴量はそれぞれ個別の全結合層に入力され，32 次元の埋め込みに変換された．次に，すべての特徴量を結合し，出力サイズが 96 次元の全結合層に入力して特徴量の融合を行った．最後に，結合された 96 次元のベクトルをさらに全結合層に入力し，9 種類の相槌ラベルの予測結果を出力を生成した．相槌ラベルを予測は，マルチラベル分類問題として扱った．評価指標は，重み付きの適合率，再現率，F 値を用いた．活性化関数として，Leaky ReLU [124] を使用し，隠れ層の値を正規化するために Layer Normalization [125] を採用した．Layer Normalization は，各サンプル内の隠れ層の値の平均と分散を基に正規化を行うものである．

本研究では次のタスクを 100 回繰り返して行い，各モデルの性能を算出した．

(1) データセットを無作為に訓練データ 90%，テストデータ 10%に分割する．このとき，訓練データとテストデータにおいて各クラスに属するシーン数の割合が同じになるように分割した．

(2) 訓練データで構築したモデルを用いて，テストデータの相槌ラベルを予測する．

表 6.5: 抽出した品詞

大分類	小分類
フィラー	フィラー
感動詞	感動詞
形容詞	形容詞
助詞	助詞（格助詞），助詞（副助詞），助詞（連体化）， 助詞（接続助詞），助詞（特殊），助詞（副詞化）， 助詞（副助詞），助詞（副助詞／並立助詞／終助詞）， 助詞（並立助詞），助詞（連体化），
助動詞	助動詞
接続詞	接続詞
動詞	動詞（自立），動詞（接尾），動詞（非自立）
名詞	名詞（サ変接続），名詞（ナイ形容詞語幹），名詞（一般）， 名詞（引用文字列），名詞（形容動詞語幹），名詞（固有名詞）， 名詞（数），名詞（接続詞的），名詞（接尾），名詞（代名詞）， 名詞（動詞非自立的），名詞（特殊），名詞（非自立）， 名詞（副詞可能）
連体詞	連体詞
副詞	副詞
接頭詞	接頭詞
記号	記号

6.3.3　機械学習モデルの性能

6.3.2 項で構築した機械学習モデルにおける予測性能を表 6.7 に示す．表 6.7 の各指標は，各モデルを 100 回繰り返し構築した際の平均値である．ベースラインモデルは，学習データにおける各ラベルの割合（表 6.6）に基づいて各ラベルを出力するモデルである．各モデル間の F 値に対して対応のない t 検定を行い，その後ボンフェローニ補正を行った．その結果，各モデル間に 1%水準で有意差が認められた．

6.3.4　相槌機能の予測に寄与するモダリティについて

本項では，相槌機能の予測に寄与したモダリティについて述べる．

6.3.4.1　ユニモーダル特徴量を用いた予測

表 6.7 より，ユニモーダル特徴量を用いて予測した場合（M1〜M3），これらのモデルはベースラインモデルよりも性能を向上させることができた．このことから，話し手の視

表 6.6: 予測対象となる各相槌ラベルの数と割合

ラベル	数	割合
N (Neutral word)	1577	32.2%
P (Positive word)	689	14.0%
NP (Non-positive word)	85	1.7%
E (Emotional word)	1171	23.9%
A (Anticipation)	540	11.0%
C (Confirmation)	379	7.7%
R (Repetition of speaker's utterance)	153	3.1%
S (Summary of speaker's utterance)	259	5.3%
O (Other)	41	0.8%

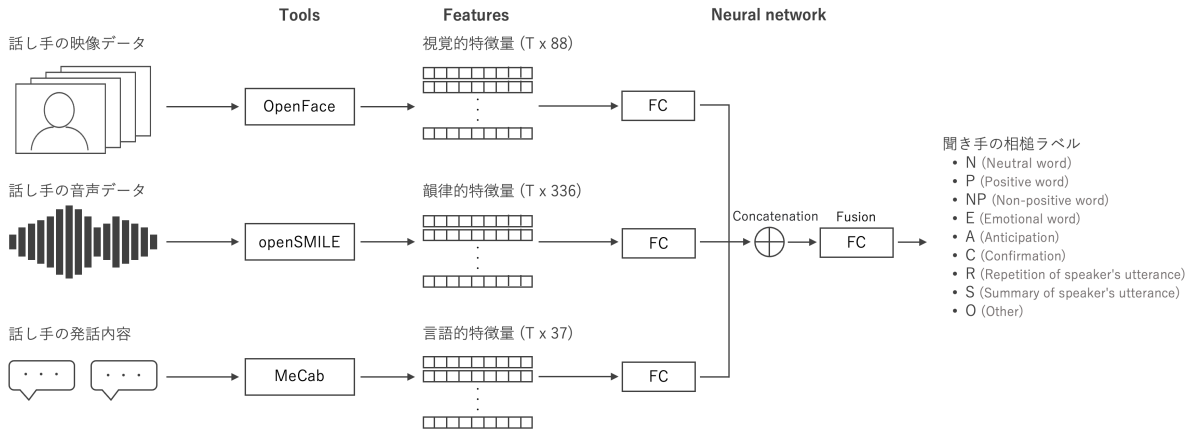

図 6.3: 相槌機能を予測する機械学習モデルの処理の流れ

覚的・韻律的・言語的特徴量から聞き手の多様な相槌機能を予測できると考えられる．その中で，最も高い予測性能を達成したのは，話し手の韻律的特徴量を用いたモデル（M2）であり，予測性能（F 値）は $F = 0.390$ であった．このモデルは，視覚的特徴量や言語的特徴量を用いたモデル（M1 と M3）と比べて予測性能（F 値）が高かった．相槌の予測や生成を行う既存研究において韻律的情報が重要であると考えられている [9]．そのため，韻律的特徴量を用いて相槌の有無やタイミングを予測する研究が多く行われている [70–78,80]．本研究の結果もこれらの研究と同じような傾向であった．聞き手の多様な相槌機能を予測するためには，韻律的特徴量を用いることが重要であると考えられる．

表 6.7: 相槌機能の予測における機械学習モデルの性能（N=100）

	視覚的	韻律的	言語的	適合率	再現率	F 値
Baseline				0.322	0.318	0.319
M1	✓			0.579	0.343	0.372
M2		✓		0.559	0.359	0.390
M3			✓	0.560	0.264	0.264
M4	✓	✓		0.597	0.389	0.422
M5	✓		✓	0.611	0.368	0.402
M6		✓	✓	0.568	0.331	0.358
M7	✓	✓	✓	**0.513**	**0.437**	**0.458**

6.3.4.2　マルチモーダル特徴量を用いた予測

表 6.7 より，マルチモーダル特徴量を用いて予測した場合（M4〜M7），これらのモデルはベースラインモデルよりも性能を向上させることができた．その中で，最も高い予測性能を達成したのは，話し手の韻律的特徴量を用いたモデル（M7）であり，予測性能（F 値）は $F = 0.458$ であった．このことから，聞き手の多様な相槌機能を予測するためには，話し手の視覚的・韻律的・言語的特徴量が有用であると考えられる．

6.3.5　相槌ラベルごとの予測性能について

最も予測性能が高かったモデル（M7）における相槌ラベルごとの予測性能を表 6.8 に示す．この結果から，ラベル N，ラベル E，ラベル A の予測性能は，表 6.8 のベースラインモデルの予測性能よりも向上していることが確認できた．一方で，ラベル P，ラベル NP，ラベル C，ラベル R，ラベル S，ラベル O の予測性能は．表 6.8 のベースラインモデルの予測性能を上回らなかった．その原因のひとつとして，表 6.6 に示すように，各相槌ラベルのデータ数が不均衡であることが挙げられる．本研究では，聞き手の相槌機能を予測する取り組みのファーストステップとして，機械学習モデルを構築する際に，各相槌ラベルのデータ数を均等にすることや，各相槌ラベルのデータ数に合わせて重み付けを行うことを取り入れなかった．今後，機械学習モデルの再構築などを行う際は，各相槌ラベルのデータ数を均等にすることや，各相槌ラベルのデータ数に合わせて重み付けを行うことを取り入れる必要がある．

6.4　本章のまとめ

本章では，話し手の言語・非言語行動から聞き手の相槌機能を予測する機械学習モデルの構築を行った．本取り組みでは実対話データと聞き手の相槌機能に関するラベルが記録

表 6.8: M7 における相槌ラベルごとの予測性能（N=100）

	適合率	再現率	F 値
N (Neutral word)	0.657	0.674	0.664
P (Positive word)	0.389	0.270	0.316
NP (Non-positive word)	0.333	0.010	0.015
E (Emotional word)	0.568	0.522	0.543
A (Anticipation)	0.469	0.345	0.395
C (Confirmation)	0.310	0.167	0.214
R (Repetition of speaker's utterance)	0.230	0.021	0.033
S (Summary of speaker's utterance)	0.240	0.094	0.131
O (Other)	0.690	0.030	0.030

された対話コーパスを用いた．この対話コーパスには，実際の 2 者対話が 13 組分記録されている．加えて，3 名の評価者が，聞き手の相槌がどのような機能であるのか判定した結果が記録されている．

話し手の言語・非言語行動から聞き手の相槌機能を予測する機械学習モデルを構築した．機械学習モデルで用いる特徴量は，話し手の視覚的，韻律的，言語的特徴量とした．話し手の視覚的，韻律的，言語的特徴量を入力とし，聞き手の相槌機能を出力するニューラルネットワークモデルを構築した．6.3.3 項より，聞き手の相槌機能を予測する機械学習モデルを構築した結果，話し手の視覚的特徴量，韻律的特徴量，言語的特徴量を用いたモデル（M7，$F = 0.458$）が，最も予測性能を向上させることができた．このことから，話し手の視覚的特徴量，韻律的特徴量，言語的特徴量に関するモダリティが，聞き手の相槌機能を予測する上で有用な振舞いであることが示唆された．

本取り組みでは上記の結果が得られたが，複数の制約が存在する．はじめに，本取り組みでは，話し手の発話終了前後の聞き手の相槌に着目するため，話し手の発話終了 0.5 秒前から 1 秒後の範囲内に存在する聞き手の発話区間に付与されている相槌ラベルを推定対象とした．そのため，話し手の発話終了 0.5 秒前から 1 秒後の範囲外に存在する相槌ラベルは，本取り組みでは扱わなかった．聞き手の相槌は，話し手が発話している最中や，話し手の発話終了から少し間をおいて行われることも想定される．今後，相槌ラベルの抽出範囲を再検討し，今回は推定対象とできなかった相槌ラベルも含めて，聞き手の相槌機能を予測できるようにする必要がある．

さらに，本取り組みでは，機械学習モデルを構築する際に，各相槌ラベルのデータ数を均等にすることや，各相槌ラベルのデータ数に合わせて重み付けを行うことを取り入れなかった．そのため，予測可能な相槌ラベルと，予測が難しい相槌ラベルが存在した．今後，機械学習モデルの再構築などを行う際は，各相槌ラベルのデータ数を均等にすることや，各相槌ラベルのデータ数に合わせて重み付けを行うことを取り入れる必要がある．

最後に，本取り組みでは，比較的次元数の少ない特徴量を用いた．そのため，各モダリティ

の特徴が，適切に抽出されていない可能性がある．今後，ResNet-50 [126]，VGGish [127]，BERT [117] のような高次元の特徴量を用いることで，予測性能が向上するか評価する必要がある．

第7章　結論

日常生活や社会活動において，人間同士のコミュニケーションは必要不可欠なものである．その中で，聞き手が話し手に対して行うフィードバックは，コミュニケーションを成立させる上で重要な要素である．しかし，人とのコミュニケーションが苦手な人は，上手にフィードバックを行えないという問題や，どのように言語・非言語行動を用いると適切なフィードバックとなるのかわからないという問題がある．本研究では，話し手および聞き手の言語・非言語行動から聞き手のフィードバックの評価や聞き手のフィードバックの種類を分類する機械学習モデルの構築を行うことで，これらの問題の解決を目指した．

本研究では，より良好な人間関係の構築が期待でき，より円滑なコミュニケーションを実現する上で重要な役割を果たす (1) 褒める行為，(2) 理解を表す行為，(3) 相槌を打つ行為に着目し，次の 3 つの課題に取り組んだ．

第 1 の課題では，聞き手および話し手の言語・非言語行動から褒める行為の検出や褒め方の上手さを推定する機械学習モデルの構築を行った．はじめに，本課題では，実対話データと褒める行為に関する評価値が記録された対話コーパスを構築した．次に，聞き手と話し手の言語・非言語行動から視覚的・韻律的・言語的特徴量を抽出し，褒める行為を検出する機械学習モデルと褒め方の上手さを推定する機械学習モデルを構築した．褒める行為を検出する機械学習モデルを構築した結果，聞き手の韻律的特徴量と発話埋め込み特徴量と話し手の視覚的特徴量と韻律的特徴量を用いたモデル（M55，$F = 0.827$）が最も検出性能を向上させることができた．このことから，聞き手の韻律的特徴量と発話埋め込み特徴量と話し手の視覚的特徴量と韻律的特徴量に関するモダリティが，対話中の褒める行為を検出する上で重要な振舞いであることが示唆された．褒め方の上手さを推定する機械学習モデルを構築した結果，聞き手の視覚的特徴量，韻律的特徴量，発話埋め込み特徴量話し手の韻律的特徴量を用いたモデル（M46，$F = 0.611$）が最も推定性能を向上させることができた．このことから，聞き手の視覚的特徴量，韻律的特徴量，発話埋め込み特徴量と話し手の韻律的特徴量に関するモダリティが，対話中の褒め方の上手さを推定する上で重要な振舞いであることが示唆された．

第 2 の課題では，聞き手の言語・非言語行動から聞き手の理解度を推定する機械学習モデルの構築を行った．本課題では，実対話データと理解度に関する評価値が記録された対話コーパスを利用し，聞き手の言語・非言語行動から視覚的・韻律的・言語的特徴量を抽出し，理解度を推定する機械学習モデルを構築した．その結果，聞き手の視覚的，韻律的，言語的特徴量を用いたモデル（M7，$F = 0.599$）が，最も推定性能を向上させることができた．このことから，聞き手の視覚的，韻律的，言語的特徴量に関するモダリティが，聞き手の理解度を推定する上で有用な振舞いであることが示唆された．

第 3 の課題では，話し手の言語・非言語行動から聞き手の相槌機能を予測する機械学習モデルの構築を行った．本課題では，実対話データと聞き手の相槌機能に関するラベルが記録された対話コーパスを利用し，話し手の言語・非言語行動から視覚的・韻律的・言語的特徴量を抽出し，聞き手の相槌機能を予測する機械学習モデルを構築した．その結果，話し手の視覚的特徴量，韻律的特徴量，言語的特徴量を用いたモデル（M7，$F = 0.458$）が，最も予測性能を向上させることができた．このことから，話し手の視覚的特徴量，韻律的特徴量，言語的特徴量に関するモダリティが，聞き手の相槌機能を予測する上で重要

な振舞いであることが示唆された．

本研究では上記の結果が得られたが，聞き手フィードバック，対話コーパス，モダリティ，機械学習モデルの観点で複数の制約が存在する．

聞き手フィードバックの制約:
本研究は，聞き手のフィードバックとして (1) 褒める行為，(2) 理解を表す行為，(3) 相槌を打つ行為に着目した．2.2.3 項で述べたように，コミュニケーション中の聞き手のフィードバックは肯定的，矯正的，否定的の評価側面に基づいて分類され，相槌，褒め，理解，共感，感謝などの肯定的なフィードバックがある．しかし，本研究は肯定的な評価側面に基づいた聞き手のフィードバックに着目したため，矯正的，否定的な評価側面に基づいた聞き手フィードバックの評価・種類を自動推定する取り組みを十分に行えていない．さらに，2.2.3 項で取り上げた肯定的なフィードバックのうち，本研究は共感や感謝に関するフィードバックの評価・種類を自動推定する取り組みを行っていない．今後，本研究で着目していない聞き手のフィードバックの評価・種類を自動推定する取り組みを行う必要がある．本研究は，(1) 褒める行為，(2) 理解を表す行為，(3) 相槌を打つ行為を分析対象としたが，褒める行為の検出，褒め方の上手さの推定，理解度の推定，相槌機能の予測はそれぞれの行為の分析に適した対話コーパスを用いた．実際のコミュニケーションでは，上記の 3 つ行為を適切に用いてフィードバックを行うことが多い．今後，(1) 褒める行為，(2) 理解を表す行為，(3) 相槌を打つ行為を一つの対話コーパスを用いて自動推定することで，より実際のコミュニケーションと同じような条件で推定できるのか検証する必要がある．

対話コーパスの制約:
本研究では，母国語が日本語の参加者を対象とした対話コーパス [97, 121] を利用した．そのため，褒める行為，理解を表す行為，相槌を打つ行為に関する判定・評価の対象や，特徴量として用いた言語・非言語行動はすべて日本人話者によるものであった．このことから，本研究で得られた知見は，日本人もしくは日本の文化に精通している人に対して有用であると考えられる．コミュニケーション中の言語・非言語行動は，文化，宗教，価値観の違いによって，同じ意味でも異なる行動を示すことがある [128]．例えば，日本では頭を縦に振り肯定を表現するが，ブルガリアでは頭を横に振り肯定を表現する．あるいは，日本では頭を横に振り否定を表現するが，ギリシャでは頭を縦に振り否定を表現する．このように，ひとつの振舞いでも，文化によって異なる意味を表している．今後，本研究で対象とした日本人とは異なる文化，宗教，価値観を持つ人々を対象とした対話コーパスを利用もしくは構築して，検証する必要がある．
本研究で利用した実対話データは，可能な限り参加者の年代を揃え，初対面同士という関係性となるようにした．その理由として，人々の関係性は複雑であり，このことがコミュニケーション中の行動や機械学習モデルの性能に影響することを最小限にするためである．日常生活や社会活動で行われるコミュニケーションの多くは，友人同士，上司と部下，教師と生徒のような人間関係のある中で行われることが多い．さらには，2 者間でコミュニケーションを行う場合だけでなく，グループでコニュニケーションを行う場合もあ

る．本研究をより発展させていくためには，様々なシーン，様々な人間関係で行われるコミュニケーションを記録し，その中で行われる聞き手のフィードバックを推定・分類できるのか明らかにする必要がある．

本研究では褒める行為の判定，褒め方の上手さの評価，理解度の評価，相槌機能のラベル付け等のアノテーションを，大学生もしくは一般の方にアノテータとして参加してもらい，判定・評価を行った．これは，アノテータの先入観や事前知識を持っていると，それらが判定・評価に影響してしまう可能性があったためである．一方で，専門家や経験者に判定・評価してもらう方が，より正確な判定・評価の結果になる可能性もある．既存研究においても，人事経験者にコミュニケーション能力について評価してもらっている研究も存在する [30]．今後，専門家や経験者に判定・評価してもらい，一般のアノテータとの結果の比較，検討を行う必要がある．

モダリティの制約:

本研究は，話し手および聞き手の視覚的，韻律的，言語的特徴量に着目し，聞き手のフィードバックの評価や聞き手のフィードバックの種類を分類する機械学習モデルの構築を行った．実際のコミュニケーションでは，話し手および聞き手のジェスチャや生体情報も円滑なコミュニケーションを行う上で重要になると考えられる．今後，ジェスチャや生体情報に関する特徴量を用いて，聞き手のフィードバックの評価や聞き手のフィードバックの種類を推定することができるのか検証する必要がある．

機械学習モデルの制約:

本研究は，話者（話し手と聞き手）が発話しているシーンから特徴量を抽出した．その理由として，コミュニケーション中の聞き手のフィードバックの評価や聞き手のフィードバックの種類を分類する機械学習モデルを構築する初期検討として，話者が発話しているシーンのみに着目したためである．そのため，話者が発話をしていないシーンは，推定や検出に使用していない．コミュニケーションでは，発話をしていなくても話者がジェスチャをしている場合や表情を変化させていることもある．今後，話者が発話をしていないシーンも対象とする必要がある．さらに，対話分析では，対話の流れを理解する必要があるため，時系列情報が重要である，今後，時系列情報を考慮した Long Short-Term Memory [129] を用いた機械学習モデルを構築することも検討する必要がある．

本研究で構築した機械学習モデルは，訓練データ 90%とテストデータ 10%に分割する方法を採用した．多くの既存研究では，Leave-one-out 交差検証 [31–33, 35, 37] や 10 分割交差検証 [29, 30, 36] を用いている．今後，機械学習モデルを改善する際や，新たな機械学習モデルを構築する際は交差検証を採用すべきである．

以上より，本研究では，コミュニケーション中の言語・非言語行動から機械学習を用いて，(1) 褒める行為の検出や褒め方の上手さの推定が可能であり，聞き手の韻律的，言語的特徴量と話し手の韻律的特徴量に関する行動が有用であること，(2) 聞き手の理解度の推定が可能であり，話し手の視覚的，韻律的特徴量と聞き手の視覚的，韻律的，言語的特徴量に関する行動が有用であること，(3) 聞き手の相槌機能を予測することが可能であり，

話し手の視覚的，韻律的，言語的特徴量が有用であることが明らかになった．本研究は，人とのコミュニケーションが苦手な人やコミュニケーションスキルが乏しい人が熟練した他者の協力を得なくても，自身のフィードバックの傾向を把握したり改善するための手がかりが得られたりするようになることを目指している．上記の目標を達成するためには，自身の言語・非言語行動からフィードバックの評価・種類を自動推定し，自身のフィードバック技能の把握や向上を促すシステムを構築する必要がある．本論文は，このシステムを実現するための要素技術に貢献する．具体的には，本論文の成果は，聞き手のフィードバックを自動的に評価する技術や，適切なフィードバックを自動的に判定して提示する技術の実現に重要な役割を果たす．本論文に基づく要素技術によってシステムが実現すれば，人とのコミュニケーションが苦手な人やコミュニケーションスキルが乏しい人が熟練した他者の協力を得なくても，自身のフィードバックの傾向を把握したり改善するための手がかりが得られたりするようになることが期待される．したがって，本研究は日常生活や社会活動で人間同士や人とコンピュータによる豊かなコミュニケーションの実現に寄与しうるものであると考えられる．

謝辞

本研究は，非常に多くの方々のご支援のもと行われました．

まず第一に，本研究の主査であり，学部時代から合わせて 7 年もの間ご指導賜りました宮田章裕教授に感謝申し上げます．私に研究の取り組み方，難しさ，面白さを教えてくださったおかげで，研究者を志した私があります．先生にご指導いただけたことで乗り越えられた壁は数多くあります．先生のような研究者を目指し，一つでも多くの壁を自分の力で乗り越えられるよう，精進して参りたいと思います．

本稿で副査を務めてくださいました尾崎知伸教授に感謝申し上げます．大変お忙しい中，研究の本質や論文の体裁の細かい部分に至るところまで非常に丁寧にご指摘いただきありがとうございます．

同じく副査を務めてくださいました北原鉄朗教授に感謝申し上げます．研究の位置付けや貢献など，本質に関わる部分を非常に丁寧にご指導いただきありがとうございます．研究以外の場面でも，いつも気にかけていただいたり，声をかけていただいたこと，とても嬉しく思います．ありがとうございます．

情報科学部門長という立場からご指導くださいました森山園子教授に感謝申し上げます．大変お忙しい中，ご丁寧にご指導・ご対応してくださったおかげで，本日に至るまで博士研究を進められていると思います．

本稿で副査を務めてくださり，共同研究者として 7 年もの間ご指導くださいました日本電信電話株式会社 NTT 人間情報研究所の石井亮特別研究員に感謝申し上げます．大変お忙しい中，いつも丁寧にご指導いただいたおかけで，数多くの研究成果を挙げることができました．私にとって，石井さんは目標としている研究者であります．石井さんと一緒に研究をしていたおかげで，現在の私があります．石井さんのように，研究分野の最前線で活躍し続けられる研究者になれるよう，精進して参りたいと思います．

様々な場面でいつも声をかけていただき，ご指導くださいました明治大学総合数理学部の小林稔専任教授に感謝申し上げます．研究に関することだけでなく，就職や将来の人生のことまで，たくさんのご相談に乗っていただいたことに大変感謝しております．

研究室の後輩であり，本研究に協力してくださった山内愛里沙さん，大串旭君，木下峻一君，田原陽平君，東直輝君，岡哲平君，鹿摩大智君に感謝申し上げます．数多くある宮田研究室の研究テーマの中から，本研究に興味を示し，私とともに本研究に取り組んでくれたことに感謝いたします．私から始めた研究テーマですが，みなさんの努力もあり，宮田研究室を代表する大きな研究テーマになりました．これからは，大串君，東君，鹿摩君が本研究テーマをより大きく発展させてくれることでしょう．期待しています．

研究室の先輩であり，私が宮田研究室に配属されてから暖かく見守ってくださった呉健朗さんに感謝申し上げます．ご自身がどんなに忙しいときでも後輩学生のことを気にかけ，丁寧に指導している姿，とても尊敬しています．これからも宮田研究室のみんなを引っ張っていく存在で居続けてください．

卒業生であり，研究室の同期である柴田万里那さんに感謝申し上げます．柴田さんの協

力があり，一生の宝物となるような対話コーパスを作成することができました．いつも私の右隣で研究を手伝ってくれていたこと，そして様々な相談相手になってくれていたことに大変感謝しております．

卒業生であり，研究室の同期である内田大樹君，小林優維さん，鈴木颯馬君，立花巧樹君，本岡宏將君に感謝申し上げます．柴田さんを含め，みなさんと一緒に宮田研究室に配属され，そしてみなさんと一緒に過ごせたこと，とても幸せに思っています．これからも長い付き合いになると思います．よろしくお願いいたします．

研究室の後輩である齊藤孝樹君，佐子柊人君，須賀美月さん，土岐田力輝君，新山はるなさんに感謝申し上げます．東君を含め，みなさんがいつも明るく，そして楽しく過ごしている様子を見ていて，私は元気をもらっていました．研究活動や日常生活の中で一筋縄ではいかないことに直面することがあると思いますが，みなさんで協力し合って乗り越えていくことを応援しています．

これまでにあげた方々のほかにも，この研究を，そして私を支えてくださった方々は数え切れません．学会で適切なアドバイスをしてくださった方々，本研究の対話収録やアノテーションに参加してくださった方々，学生時代を共に過ごしてくれた友人達，そして家族に感謝の意を表します．

参考文献

[1] 深田博己. インターパーソナル・コミュニケーション：対人コミュニケーションの心理学. 北大路書房, 1998.

[2] Joseph A. DeVito. *The Interpersonal Communication Book*. Harper and Row, 1986.

[3] Akira Morikawa, Ryo Ishii, Hajime Noto, Atsushi Fukayama, and Takao Nakamura. Determining most suitable listener backchannel type for speaker's utterance. In *Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents (IVA '22)*, pp. 1–3, 2022.

[4] James C. McCroskey and Virginia P. Richmond. *Fundamentals of human communication: an interpersonal perspective*, pp. 169–214. Waveland Press Inc., 1996.

[5] Gary R. VandenBos. APA 心理学大辞典. 培風館, 2013.

[6] Herbert H. Clark. *Using language*. Cambridge University Press, 1996.

[7] Herbert H. Clark and Meredyth A. Krych. Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, Vol. 50, No. 1, pp. 62–81, 2004.

[8] Emanuel A. Schegloff. *Discourse as an Interactional Achievement: Some Uses of 'uh huh' and Other Things That Come Between Sentences*, pp. 71–93. Georgetown University Press, 1982.

[9] Victor H. Yngve. *On Getting a Word in Edgewise*, pp. 567–578. Chicago Linguistic Society, 1970.

[10] Maurice Piéron and John Cheffers. *Research in Sport Pedagogy: Empirical Analytical Perspective*. Hofmann, 1988.

[11] William S. Verplanck. The control of the content of conversation: reinforcement of statements of opinion. *Journal of Abnormal and Social Psychology*, Vol. 51, No. 3, pp. 668–676, 1955.

[12] Jens Allwood, Joakim Nivre, and Elisabeth Ahlsén. On the semantics and pragmatics of linguistic feedback. *Journal of semantics*, Vol. 9, No. 1, pp. 1–26, 1992.

[13] Janet B. Bavelas, Linda Coates, and Trudy Johnson. Listeners as co-narrators. *Journal of personality and social psychology*, Vol. 79, No. 6, pp. 941–952, 2000.

[14] John M. Wiemann and Mark L. Knapp. Turn-taking in conversations. *Journal of Communication*, Vol. 25, No. 2, pp. 75–92, 1975.

[15] Adam Kendon. Some functions of gaze-direction in social interaction. *Acta Psychologica*, Vol. 26, pp. 22–63, 1967.

[16] Robert E. Kraut, Steven H. Lewis, and Lawrence W. Swezey. Listener responsiveness and the coordination of conversation. *Journal of personality and social psychology*, Vol. 43, No. 4, pp. 718–731, 1982.

[17] Starkey Duncan. On the structure of speaker–auditor interaction during speaking turns. *Language in society*, Vol. 3, No. 2, pp. 161–180, 1974.

[18] Allen T. Dittmann. *The Body Movement-speech Rhythm Relationship as a Cue to Speech Encoding*, pp. 135–152. Elsevier, 1972.

[19] Howard M. Rosenfeld. The experimental analysis of interpersonal influence processes. *Journal of Communication*, Vol. 22, No. 4, pp. 424–442, 1972.

[20] Howard M. Rosenfeld. *Conversational Control Functions of Nonverbal Behavior*, pp. 563–601. Psychology Press, 1978.

[21] Allen T. Dittmann and Lynn G. Llewellyn. Relationship between vocalizations and head nods as listener responses. *Journal of Personality and Social Psychology*, Vol. 9, No. 1, pp. 79–84, 1968.

[22] Matthew Turk. Multimodal interaction: A review. *Pattern Recognition Letters*, Vol. 36, pp. 189–195, 2014.

[23] Alessandro Vinciarelli, Maja Pantic, and Hervé Bourlard. Social signal processing: Survey of an emerging domain. *Image and Vision Computing*, Vol. 27, No. 12, pp. 1743–1759, 2009.

[24] Judee K. Burgoon, Nadia Magnenat-Thalmann, Maja Pantic, and Alessandro Vinciarelli. *Social Signal Processing*. Cambridge University Press, 2017.

[25] 東中竜一郎, 岡田将吾, 藤江真也, 森大毅. 対話システムと感情. 人工知能, Vol. 31, No. 5, pp. 664–670, 2016.

[26] Ligia Batrinca, Nadia Mana, Bruno Lepri, Nicu Sebe, and Fabio Pianesi. Multimodal personality recognition in collaborative goal-oriented tasks. *IEEE Transactions on Multimedia*, Vol. 18, No. 4, pp. 659–673, 2016.

[27] Fabio Valente, Samuel Kim, and Petr Motlicek. Annotation and recognition of personality traits in spoken conversations from the ami meetings corpus. In *Proceedings of the 13th Annual Conference of the International Speech Communication Association (INTERSPEECH '12)*, pp. 1183–1186, 2012.

[28] Dineshbabu Jayagopi, Dairazalia Sanchez-Cortes, Kazuhiro Otsuka, Junji Yamato, and Daniel Gatica-Perez. Linking speaking and looking behavior patterns with group composition, perception, and performance. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction (ICMI '12)*, pp. 433–440, 2012.

[29] Sunghyun Park, Han Suk Shim, Moitreya Chatterjee, Kenji Sagae, and Louis-Philippe Morency. Computational analysis of persuasiveness in social multimedia: A novel dataset and multimodal prediction approach. In *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14)*, pp. 50–57, 2014.

[30] Shogo Okada, Yoshihiko Ohtake, Yukiko I. Nakano, Yuki Hayashi, Hung-Hsung Huang, Yutaka Takase, and Katsumi Nitta. Estimating communication skills using dialogue acts and nonverbal features in multiple discussion datasets. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction (ICMI '16)*, pp. 169–176, 2016.

[31] Vikram Ramanarayanan, Chee Wee Leong, Lei Chen, Gary Feng, and David Suendermann-Oeft. Evaluating speech, face, emotion and body movement time-series features for automated multimodal presentation scoring. In *Proceedings of the 17th ACM on International Conference on Multimodal Interaction (ICMI '15)*, pp. 23–30, 2015.

[32] Lei Chen, Gary Feng, Jilliam Joe, Chee Wee Leong, Christopher Kitchen, and Chong Min Lee. Towards automated assessment of public speaking skills using multimodal cues. In *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14)*, pp. 200–203, 2014.

[33] Yutaro Yagi, Shogo Okada, Shota Shiobara, and Sota Sugimura. Predicting multimodal presentation skills based on instance weighting domain adaptation. *Journal on Multimodal User Interfaces*, Vol. 16, pp. 1–16, 2021.

[34] Dairazalia Sanchez-Cortes, Oya Aran, Marianne Schmid Mast, and Daniel Gatica-Perez. A nonverbal behavior approach to identify emergent leaders in small groups. *IEEE Transactions Multimedia*, Vol. 14, No. 3, pp. 816–832, 2012.

[35] Laurent Son Nguyen, Denise Frauendorfer, Marianne Schmid Mast, and Daniel Gatica-Perez. Hire me: Computational inference of hirability in employment inter-

views based on nonverbal behavior. *IEEE Transactions Multimedia*, Vol. 16, No. 4, pp. 1018–1031, 2014.

[36] Ryo Ishii, Kazuhiro Otsuka, Shiro Kumano, Ryuichiro Higashinaka, and Junji Tomita. Analyzing gaze behavior and dialogue act during turn-taking for estimating empathy skill level. In *Proceedings of the 20th ACM International Conference on Multimodal Interaction (ICMI '18)*, pp. 31–39, 2018.

[37] Zhi-Xuan Tan, Arushi Goel, Thanh-Son Nguyen, and Desmond C. Ong. A multimodal lstm for predicting listener empathic responses over time. In *Proceedings of the 14th IEEE International Conferenceon Automatic Face and Gesture Recognition (FG '19)*, pp. 1–4, 2019.

[38] Mohammad Soleymani, Kalin Stefanov, Sin-Hwa Kang, Jan Ondras, and Jonathan Gratch. Multimodal analysis and estimation of intimate self-disclosure. In *Proceedings of the 21st ACM International Conference on Multimodal Interaction (ICMI '19)*, pp. 59–68, 2019.

[39] Senko K. Maynard. On back-channel behavior in japanese and english casual conversation. *Linguistics*, Vol. 24, No. 6, pp. 1079–1108, 1986.

[40] Jere Brophy. Teacher praise: A functional analysis. *Review of educational research*, Vol. 51, No. 1, pp. 5–32, 1981.

[41] Jennifer Henderlong and Mark R. Lepper. The effects of praise on children's intrinsic motivation: A review and synthesis. *Psychological Bulletin*, Vol. 128, No. 5, pp. 774–795, 2002.

[42] Albert Mehrabian and Norman Epstein. A measure of emotional empathy. *Journal of personality*, Vol. 40, pp. 525–543, 1972.

[43] Jean Decety and Margarita Svetlova. Putting together phylogenetic and ontogenetic perspectives on empathy. *Developmental cognitive neuroscience*, Vol. 2, No. 1, pp. 1–24, 2012.

[44] Jamil Zaki and Kevin N. Ochsner. The neuroscience of empathy: progress, pitfalls and promise. *Nature neuroscience*, Vol. 15, No. 5, pp. 675–680, 2012.

[45] Michael E. McCullough, Shelley D. Kilpatrick, Robert A. Emmons, and David B. Larson. Is gratitude a moral affect? *Psychological Bulletin*, Vol. 127, No. 2, pp. 249–266, 2001.

[46] Tara M. Kalis, Kimberly J. Vannest, and Rich Parker. Praise counts: Using self-monitoring to increase effective teaching practices. *Preventing School Failure: Alternative Education for Children and Youth*, Vol. 51, No. 3, pp. 20–27, 2007.

[47] Lyndsay N. Jenkins, Margaret T. Floress, and Wendy Reinke. Rates and types of teacher praise: A review and future directions. *Psychology in the Schools*, Vol. 52, No. 5, pp. 463–476, 2015.

[48] Nigel Ward. Pragmatic functions of prosodic features in non-lexical utterances. In *Proceedings of the 2nd International Conference on Speech Prosody (SP '04)*, pp. 325–328, 2004.

[49] Kazuhiro Otsuka and Masahiro Tsumori. Analyzing multifunctionality of head movements in face-to-face conversations using deep convolutional neural networks. *IEEE Access*, Vol. 8, No. 3, pp. 217169–217195, 2020.

[50] 飯塚海斗, 大塚和弘. 対話中における聞き手の頭部運動と相槌の相乗機能の解析. 人工知能学会論文誌, Vol. 38, No. 3, pp. J–M91_1–17, 2023.

[51] 大西俊輝, 山内愛里沙, 大串旭, 石井亮, 青野裕司, 宮田章裕. 褒める行為における頭部・顔部の振舞いの分析. 情報処理学会論文誌, Vol. 62, No. 9, pp. 1620–1628, 2021.

[52] Toshiki Onishi, Arisa Yamauchi, Asahi Ogushi, Ryo Ishii, Atsushi Fukayama, Takao Nakamura, and Akihiro Miyata. Modeling japanese praising behavior by analyzing audio and visual behaviors. *Frontiers in Computer Science*, Vol. 4, , 2022.

[53] Toshiki Onishi, Asahi Ogushi, Ryo Ishii, Atsushi Fukayama, and Akihiro Miyata. Prediction of praising skills based on multimodal information. In *Proceedings of the 12th International Conference on Affective Computing and Intelligent Interaction (ACII '24)*, pp. 176–184, 2024.

[54] Shunichi Kinoshita, Toshiki Onishi, Naoki Azuma, Ryo Ishii, Atsushi Fukayama, Takao Nakamura, and Akihiro Miyata. A study of prediction of listener's comprehension based on multimodal information. In *Proceedings of the the 23rd ACM International Conference on Intelligent Virtual Agents (IVA '23)*, No. 30, pp. 1–4, 2023.

[55] Toshiki Onishi, Naoki Azuma, Shunichi Kinoshita, Ryo Ishii, Atsushi Fukayama, Takao Nakamura, and Akihiro Miyata. Prediction of various backchannel utterances based on multimodal information. In *Proceedings of the the 23rd ACM International Conference on Intelligent Virtual Agents (IVA '23)*, No. 47, pp. 1–4, 2023.

[56] James Benjamin. コミュニケーション：話すことと聞くことを中心に. 二瓶社, 1990.

[57] 石井敏. コミュニケーション研究の意義と理論背景, pp. 3–24. 桐原書店, 1993.

[58] James C. McCroskey. *An Introduction to Rhetorical Communication*. Allyn and Bacon, 2001.

[59] James C. McCroskey and Virginia P. Richmond. 非言語行動の心理学. 北大路書房, 2006.

[60] Marjorie F. Vargas. 非言語コミュニケーション. 新潮社, 1987.

[61] Charles Goodwin. Between and within: Alternative sequential treatments of continuers and assessments. *Human Studies*, Vol. 9, pp. 205–217, 1986.

[62] Starkey Duncan and Donald W. Fiske. *Face-to-face interaction: Research, methods, and theory*. Erlbaum, 1977.

[63] Sung H. Cheon, Johnmarshall Reeve, Tae H. Yu, and Hue R. Jang. The teacher benefits from giving autonomy support during physical education instruction. *Journal of Sport and Exercise Psychology*, Vol. 36, No. 4, pp. 331–346, 2014.

[64] Bruce P. Doré, Robert R. Morris, Daisy A. Burr, Rosalind W. Picard, and Kevin N. Ochsner. Helping others regulate emotion predicts increased regulation of one's own emotions and decreased symptoms of depression. *Personality and Social Psychology Bulletin*, Vol. 43, No. 5, pp. 729–739, 2017.

[65] Michael W. Beets, Kenneth H. Pitetti, and Loretta Forlaw. The role of self-efficacy and referent specific social support in promoting rural adolescent girls' physical activity. *American journal of health behavior*, Vol. 31, No. 3, pp. 227–237, 2007.

[66] Kyosuke Kakinuma, Mai Nakai, Yuki Hada, Mari Kizawa, and Ayumi Tanaka. Praise affects the "praiser": Effects of ability-focused vs. effort-focused praise on motivation. *The Journal of Experimental Education*, Vol. 90, No. 3, pp. 634–655, 2022.

[67] Robin P. Ennis, David J. Royer, Kathleen L. Lane, Holly M. Menzies, Wendy P. Oakes, and Liane E. Schellman. Behavior-specific praise: An effective, efficient, low-intensity strategy to support student success. *Beyond Behavior*, Vol. 27, No. 3, pp. 134–139, 2018.

[68] Rosemarie Anderson, Sam T. Manoogian, and J. Steven Reznick. The undermining and enhancing of intrinsic motivation in preschool children. *Journal of personality and social psychology*, Vol. 34, No. 5, pp. 915–922, 1976.

[69] Deborah Stipek, Sandro Recchia, and Scott M. McClintic. Study2: 2-5-year-olds' reactions to success and failure and the effects of praise. *Monographs of the Society for Research in Child Development*, Vol. 57, pp. 39–59, 1992.

[70] Shinya Fujie, Kenta Fukushima, and Tetsunori Kobayashi. Back-channel feedback generation using linguistic and nonlinguistic information and its application to spoken dialogue system. In *Proceedings of the 9th European Conference on Speech Communication and Technology (INTERSPEECH '05)*, pp. 889–892, 2005.

[71] Kohei Hara, Koji Inoue, Katsuya Takanashi, and Tatsuya Kawahara. Prediction of turn-taking using multitask learning with prediction of backchannels and fillers. In *Proceedings of the 19th Annual Conference of the International Speech Communication Association (INTERSPEECH '18)*, pp. 991–995, 2018.

[72] Ryo Ishii, Xutong Ren, Michal Muszynski, and Louis-Philippe Morency. Multimodal and multitask approach to listener's backchannel prediction: Can prediction of turn-changing and turn-management willingness improve backchannel modeling? In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents (IVA '21)*, pp. 131–138, 2021.

[73] Louis-Philippe Morency, Iwan De Kok, and Jonathan Gratch. Predicting listener backchannels: A probabilistic multimodal approach. In *Proceedings of the 8th International Workshop on Intelligent Virtual Agents (IVA '08)*, pp. 176–190, 2008.

[74] Markus Mueller, David Leuschner, Lars Briem, Maria Schmidt, Kevin Kilgour, Sebastian Stueker, and Alex Waibel. Using neural networks for data-driven backchannel prediction: A survey on input features and training techniques. In *International Conference on Human-Computer Interaction (HCI '15)*, pp. 329–340, 2015.

[75] Khiet P. Truong, Ronald Poppe, and Dirk Heylen. Arule-based backchannel prediction model using pitch and pause information. In *11th Annual Conference of the International Speech Communication Association (INTERSPEECH '10)*, pp. 3058–3061, 2010.

[76] Nigel Ward. Using prosodic clues to decide when to produce back-channel utterances. In *Proceedings of the 4th International Conference on Spoken Language Processing (ICSLP '96)*, Vol. 3, pp. 1728–1731, 1996.

[77] Nigel Ward and Wataru Tsukahara. Prosodic features which cue back-channel responses in english and japanese. *Journal of Pragmatics*, Vol. 32, No. 8, pp. 1177–1207, 2000.

[78] Robin Ruede, Markus Müller, Sebastian Stüker, and Alex Waibel. Yeah, right, uh-huh: A deep learning backchannel predictor. *Advanced Social Interaction with Agents*, Vol. 510, pp. 247–258, 2019.

[79] Lixing Huang, Louis-Phillippe Morency, and Jonathan Gratch. Parasocial consensus sampling: Combining multiple perspectives to learn virtual human behavior. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS '10)*, pp. 1265–1272, 2010.

[80] Auriane Boudin, Roxane Bertrand, Stéphane Rauzy, Magalie Ochs, and Philippe Blache. A multimodal model for predicting feedback position and type during conversation. *Speech Communication*, Vol. 159, p. 103066, 2024.

[81] Anne H. Anderson, Miles Bader, Ellen G. Bard, Elizabeth Boyle, Gwyneth Doherty, Simon Garrod, Stephen Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Catherine Sotillo, Henry S. Thompson, and Regina Weinert. The hcrc map task corpus. *Language and speech*, Vol. 34, No. 4, pp. 351–366, 1991.

[82] Jean Carletta, Simone Ashby, Sebastien Bourban, Mike Flynn, Mael Guillemot, Thomas Hain, Jaroslav Kadlec, Vasilis Karaiskos, Wessel Kraaij, Melissa Kronenthal, Guillaume Lathoud, Mike Lincoln, Agnes Lisowska, Iain McCowan, Wilfried Post, Dennis Reidsma, and Pierre Wellner. The ami meeting corpus: A preannouncement. In *International workshop on machine learning for multimodal interaction*, pp. 28–39, 2005.

[83] Jill Kickul and George Neuman. Emergent leadership behaviors: The function of personality and cognitive ability in determining teamwork performance and ksas. *Journal of Business and Psychology*, Vol. 15, No. 1, pp. 27–51, 2000.

[84] Paul Ekman and Wallace V. Friesen. Manual for the facial action coding system. *Palo Alto: Consulting Psychologists Press*, 1977.

[85] Fumio Nihei, Yukiko I. Nakano, Yuki Hayashi, Hung-Hsuan Huang, and Shogo Okada. Predicting influential statements in group discussions using speech and head motion information. In *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14)*, pp. 136–143, 2014.

[86] Michael A. McDaniel, Deborah L. Whetzel, Frank L. Schmidt, and Steven D. Maurer. The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of applied psychology*, Vol. 79, No. 4, pp. 599–616, 1994.

[87] David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, Gale Lucas,

Stacy Marsella, Fabrizio Morbini, Angela Nazarian, Stefan Scherer, Giota Stratou, Apar Suri, David Traum, Rachel Wood, Yuyu Xu, Albert Rizzo, and Louis-Philippe Morency. Simsensei kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems (AAMAS '14)*, pp. 1061–1068, 2014.

[88] Sin-Hwa Kang and Jonathan Gratch. Virtual humans elicit socially anxious interactants' verbal self - disclosure. *Computer Animation and Virtual Worlds*, Vol. 21, No. 3–4, pp. 473–482, 2010.

[89] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.

[90] Fiona J. Buckingham, Keeley A. Crockett, Zuhair A. Bandar, James D. O'Shea, Kathleen. M. MacQueen, and Mario Chen. Measuring human comprehension from nonverbal behaviour using artificial neural networks. In *Proceedings of the 2012 International Joint Conference on Neural Networks (IJCNN '12)*, pp. 1–8, 2012.

[91] Tatsuya Kawahara, Soichiro Hayashi, and Katsuya Takanashi. Estimation of interest and comprehension level of audience through multi-modal behaviors in poster conversations. In *Proceedings of the 14th Annual Conference of the International Speech Communication Association (INTERSPEECH '13)*, pp. 1882–1885, 2013.

[92] Olcay Türk, Stefan Lazarov, Yu Wang, Hendrik Buschmeier, Angela Grimminger, and Petra Wagner. Predictability of understanding in explanatory interactions based on multimodal cues. In *Proceedings of the 26th International Conference on Multimodal Interaction (ICMI '24)*, pp. 449–458, 2024.

[93] Rod Gardner. *When listeners talk: Response tokens and listener stance*. John Benjamins, 2001.

[94] Yasuharu Den, Nao Yoshida, Katsuya Takanashi, and Hanae Koiso. Annotation of japanese response tokens and preliminary analysis on their distribution in three-party conversations. In *Proceedings of the 2011 International Conference on Speech Database and Assessments (Oriental COCOSDA '11)*, pp. 168–173, 2011.

[95] Peggy P. Hester, Jo M. Hendrickson, and Robert A. Gable. Forty years later–the value of praise, ignoring, and rules for preschoolers at risk for behavior disorders. *Education and Treatment of Children*, Vol. 32, No. 4, pp. 513–535, 2009.

[96] Eddie Brummelman, Sander Thomaes, Geertjan Overbeek, B. Orobio de Castro, Marcel A. van den Hout, and Brad J. Bushman. On feeding those hungry for praise:

Person praise backfires in children with low self-esteem. *Journal of Experimental Psychology: General*, Vol. 143, No. 1, pp. 9–14, 2014.

[97] Toshiki Onishi, Asahi Ogushi, Yohei Tahara, Ryo Ishii, Atsushi Fukayama, Takao Nakamura, and Akihiro Miyata. A comparison of praising skills in face-to-face and remote dialogues. In *Proceedings of the 13th Language Resources and Evaluation Conference (LREC '22)*, pp. 5805–5812, 2022.

[98] Hennie Brugman and Albert Russel. Annotating multimedia / multi-modal resources with elan. In *Proceedings of the 4th International Conference on Language Resources and Language Evaluation (LREC '04)*, pp. 2065–2068, 2004.

[99] Hanae Koiso, Yasuo Horiuchi, Syun Tutiya, Akira Ichikawa, and Yasuharu Den. An analysis of turn-taking and backchannels based on prosodic and syntactic features in japanese map task dialogs. *Language and Speech*, Vol. 41, pp. 295–321, 1998.

[100] Hiroki Mori, Tomoyuki Satake, Makoto Nakamura, and Hideki Kasuya. Constructing a spoken dialogue corpus for studying paralinguistic information in expressive conversation and analyzing its statistical/acoustic characteristics. *Speech Communication*, Vol. 53, No. 1, pp. 36–50, 2011.

[101] 伝康晴, 小木曽智信, 小椋秀樹, 山田篤, 峯松信明, 内元清貴, 小磯花絵. コーパス日本語学のための言語資源：形態素解析用電子化辞書の開発とその応用. 日本語科学, Vol. 22, pp. 101–123, 2007.

[102] Kazunori Komatani, Ryu Takeda, and Shogo Okada. Analyzing differences in subjective annotations by participants and third-party annotators in multimodal dialogue corpus. In *Proceedings of the 24th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL '23)*, pp. 104–113, 2023.

[103] Ryo Ishii, Xutong Ren, Michal Muszynski, and Louis-Philippe Morency. Trimodal prediction of speaking and listening willingness to help improve turn-changing modeling. *Frontiers in Psychology*, Vol. 13, , 2022.

[104] Laurence Devillers, Laurence Vidrascu, and Lori Lamel. Challenges in real-life emotion annotation and machine learning based detection. *Neural Networks*, Vol. 18, No. 42, pp. 407–422, 2005.

[105] Carlos Busso, Murtaza Bulut, Chi-Chun Lee, Abe Kazemzadeh, Emily Mower, Samuel Kim, Jeannette N. Chang, Sungbok Lee, and Shrikanth S. Narayanan. Iemocap: Interactive emotional dyadic motion capture database. *Lang Resources and Evaluation*, Vol. 42, No. 4, pp. 335–359, 2008.

[106] Dennis Reidsma and Rieks op den Akker. Exploiting 'subjective' annotations. In *Proceedings of the Workshop on Human Judgements in Computational Linguistics (COLING '08)*, pp. 8–16, 2008.

[107] Shiro Kumano, Kazuhiro Otsuka, Dan Mikami, Masafumi Matsuda, and Junji Yamato. Analyzing interpersonal empathy via collective impressions. *IEEE Transactions on Affective Computing*, Vol. 6, No. 4, pp. 324–336, 2015.

[108] Joseph L. Fleiss. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, Vol. 76, No. 5, pp. 378–382, 1971.

[109] Patrick E. Shrout and Joseph L. Fleiss. Intraclass correlations: Uses in assessing rater reliability. *Psychological bulletin*, Vol. 86, No. 2, pp. 420–428, 1979.

[110] Terry K. Koo and Mae Y. Li. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, Vol. 15, No. 2, pp. 155–163, 2016.

[111] Paul Ekman. Movements with precise meanings. *Journal of communication*, Vol. 26, No. 3, pp. 14–26, 1976.

[112] Tadas Baltrušaitis, Peter Robinson, and Louis-Philippe Morency. Openface: An open source facial behavior analysis toolkit. In *IEEE Winter Conference on Applications of Computer Vision (WACV '16)*, pp. 1–10, 2016.

[113] George L. Trager. Paralanguage: A first approximation. *Studies in Linguistics*, Vol. 13, pp. 1–12, 1958.

[114] Florian Eyben, Martin Wöllmer, and Börn Schuller. opensmile – the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM international conference on Multimedia (MM '10)*, pp. 1459–1462, 2010.

[115] Björn Schuller, Stefan Steidl, and Anton Batliner. The interspeech 2009 emotion challenge. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association (INTERSPEECH '09)*, pp. 312–315, 2009.

[116] Taku Kudo. *MeCab: yet another part-of-speech and morphological analyzer*, 2006. https://taku910.github.io/mecab/.

[117] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT '19)*, pp. 4171–4186, 2019.

[118] Leo Breiman. Random forests. *Machine Learning*, Vol. 45, No. 1, pp. 5–32, 2001.

[119] James Bergstra, Dan Yamins, and David D. Cox. Hyperopt: A python library for optimizing the hyperparameters of machine learning algorithms. In *Proceedings of the 12th Python in Science Conferences (SciPy '13)*, pp. 13–20, 2013.

[120] Takashi Yamaguchi, Koji Inoue, Koichiro Yoshino, Katsuya Takanashi, Nigel G. Ward, and Tatsuya Kawahara. Analysis and prediction of morphological patterns of backchannels for attentive listening agents. In *Proceedings of the 7th International Workshop on Spoken Dialogue Systems (IWSDS '16)*, pp. 1–12, 2016.

[121] Ryo Ishii, Ryuichiro Higashinaka, and Junji Tomita. Predicting nods by using dialogue acts in dialogue. In *Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC '18)*, pp. 2940–2944, 2018.

[122] Keith Curtis, Gareth J.F. Jones, and Nick Campbell. Effects of good speaking techniques on audience engagement. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction (ICMI '15)*, pp. 35–42, 2015.

[123] Sidney Siegel and N. John Castellan Jr. *Nonparametric Statistics for the Behavioral Sciences (2nd ed.)*. Mcgraw-Hill Book Company, 1988.

[124] Andrew L. Maas, Awni Y. Hannun, and Andrew Y. Ng. Rectifier nonlinearities improve neural network acoustic models. In *ICML Workshop on Deep Learning for Audio, Speech, and Language Processing (WDLASL '13)*, 2013.

[125] Jimmy L. Ba, Jamie R. Kiros, and Geoffrey E. Hinton. Layer normalization. *arXiv preprint arXiv:1607.06450*, 2016.

[126] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the 29th IEEE conference on computer vision and pattern recognition (CVPR '16)*, pp. 770–778, 2016.

[127] Shawn Hershey, Sourish Chaudhuri, Daniel P.W. Ellis, Jort F. Gemmeke, Aren Jansen, R. Channing Moore, Manoj Plakal, Devin Platt, Rif A. Saurous, Bryan Seybold, Malcolm Slaney, Ron J. Weiss, and Kevin Wilson. Cnn architectures for large-scale audio classification. In *The 42nd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '17)*, pp. 131–135, 2017.

[128] Jeanne M. Brett. Culture and negotiation. *International journal of psychology*, Vol. 35, No. 2, pp. 97–104, 2000.

[129] Sepp Hochreiter and Jürgen Schmidhuber. Long-short term memory. *Neural computation*, Vol. 9, No. 8, pp. 1735–1780, 1997.

付録

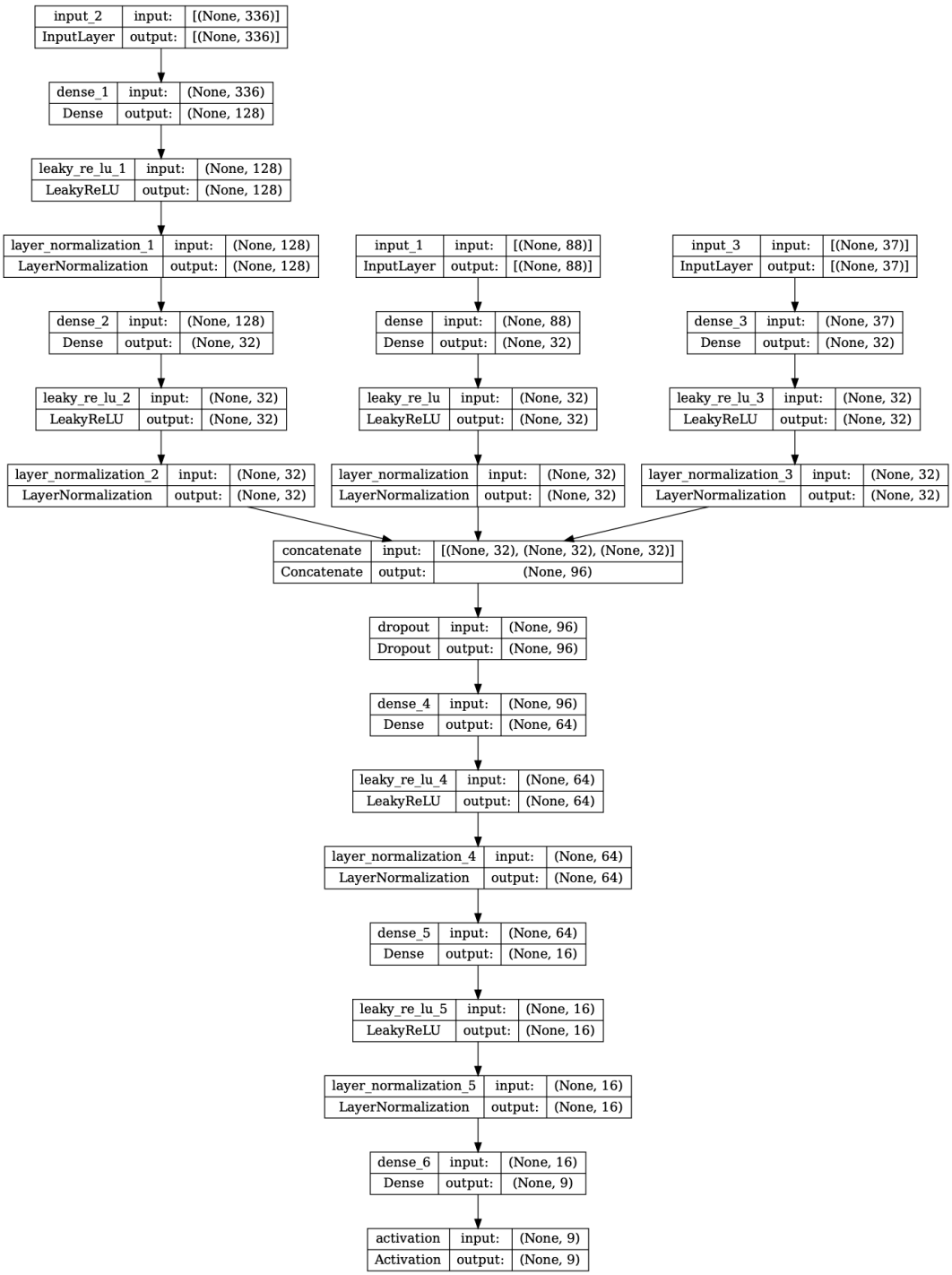

図 A.1: 6.3.2 項で構築した相槌機能を予測するモデルの全体図

研究業績

本研究に関する業績

査読付き論文誌

(1) Toshiki Onishi, Asahi Ogushi, Ryo Ishii, and Akihiro Miyata: Detecting Praising Behaviors Based on Multimodal Information. The IEICE Transactions on Information and Systems (Jan. 2025).

(2) Toshiki Onishi, Asahi Ogushi, Shunichi Kinoshita, Ryo Ishii, Atsushi Fukayama, and Akihiro Miyata: Detecting Praising Behavior Based on Multimodal Information in Remote Dialogue, Journal of Information Processing, Vol.33 (Jan. 2025).

(3) Asahi Ogushi, Toshiki Onishi, Ryo Ishii, Atsushi Fukayama, and Akihiro Miyata: Predicting Praising Skills Using Multimodal Information in Remote Dialogue. Transactions of the Virtual Reality Society of Japan, Vol.29, No.3, pp.127–138 (Sep. 2024).

(4) Toshiki Onishi, Arisa Yamauchi, Asahi Ogushi, Ryo Ishii, Atsushi Fukayama, Takao Nakamura, and Akihiro Miyata: Modeling Japanese Praising Behavior by Analyzing Audio and Visual Behaviors. Frontiers in Computer Science, Vol.4 (Feb. 2022).

(5) 大西俊輝, 山内愛里沙, 大串旭, 石井亮, 青野裕司, 宮田章裕: 褒める行為における頭部・顔部の振舞いの分析, 情報処理学会論文誌, Vol.62, No.9, pp.1620–1628 (2021 年 9 月).

査読付き国際会議

(1) Toshiki Onishi, Asahi Ogushi, Ryo Ishii, Atsushi Fukayama, and Akihiro Miyata: Prediction of Praising Skills Based on Multimodal Information. Proceedings of the 12th International Conference on Affective Computing and Intelligent Interaction (ACII 2024), pp.176–184 (Sep. 2024).

(2) Shunichi Kinoshita, Toshiki Onishi, Naoki Azuma, Ryo Ishii, Atsushi Fukayama, Takao Nakamura, and Akihiro Miyata: A Study of Prediction of Listener's Comprehension Based on Multimodal Information. Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents (IVA 2023), No.30, pp.1–4 (Sep. 2023).

(3) Toshiki Onishi, Naoki Azuma, Shunichi Kinoshita, Ryo Ishii, Atsushi Fukayama, Takao Nakamura, and Akihiro Miyata: Prediction of Various Backchannel Utterances Based on Multimodal Information. Proceedings of the 23rd ACM Interna-

tional Conference on Intelligent Virtual Agents (IVA 2023), No.47, pp.1–4 (Sep. 2023).

(4) Asahi Ogushi, Toshiki Onishi, Yohei Tahara, Ryo Ishii, Atsushi Fukayama, Takao Nakamura, and Akihiro Miyata: Analysis of Praising Skills Focusing on Utterance Contents. Proceedings of the 23rd Annual Conference of the International Speech Communication Association (INTERSPEECH 2022), pp.2743–2747 (Sep. 2022).

(5) Toshiki Onishi, Asahi Ogushi, Yohei Tahara, Ryo Ishii, Atsushi Fukayama, Takao Nakamura, and Akihiro Miyata: A Comparison of Praising Skills in Face-to-Face and Remote Dialogues. Proceedings of the 13th Language Resources and Evaluation Conference (LREC 2022), pp.5805–5812 (Jun. 2022).

(6) Toshiki Onishi, Arisa Yamauchi, Ryo Ishii, Yushi Aono, and Akihiro Miyata: Analyzing Nonverbal Behaviors along with Praising. Proceedings of the 22nd ACM International Conference on Multimodal Interaction (ICMI 2020), pp.609–613 (Oct. 2020).

査読付き国内会議

(1) 田原陽平, 大西俊輝, 大串旭, 石井亮, 深山篤, 中村高雄, 宮田章裕: 対面・遠隔対話からの称賛行為検出の基礎検討, 情報処理学会グループウェアとネットワークサービスワークショップ 2022 論文集, Vol.2022, pp.36–43 (2022 年 11 月).

(2) 大串旭, 大西俊輝, 田原陽平, 石井亮, 深山篤, 中村高雄, 宮田章裕: 言語特徴に着目した褒め方の上手さの推定モデルの検討, 情報処理学会グループウェアとネットワークサービスワークショップ 2021 論文集, Vol.2021, pp.1–8 (2021 年 11 月).

(3) 大西俊輝, 柴田万里那, 山内愛里沙, 呉健朗, 石井亮, 富田準二, 宮田章裕: 褒め方の上手さの推定における頭部・顔部の効果, 情報処理学会グループウェアとネットワークサービスワークショップ 2019 論文集, Vol.2019, pp.1–6 (2019 年 11 月).

研究会・シンポジウム

(1) 鹿摩大智, 岡哲平, 大西俊輝, 東直輝, 石井亮, 深山篤, 宮田章裕: マルチモーダル情報に基づいて適切な相槌を提示するシステムの検討, 情報処理学会シンポジウム論文集, マルチメディア、分散、協調とモバイル (DICOMO 2024), Vol.2024, pp.468–474 (2024 年 6 月).

(2) 大串旭, 大西俊輝, 石井亮, 深山篤, 宮田章裕: 表情特徴量に着目した褒め方の上手さフィードバックシステムの基礎検討, 情報処理学会インタラクション 2024 論文集, pp.653–657 (2024 年 3 月).

(3) 鹿摩大智, 岡哲平, 大西俊輝, 東直輝, 石井亮, 深山篤, 宮田章裕: 聞き手の相槌種類に応じた表情生成システムの基礎検討, 情報処理学会インタラクション 2024 論文集, pp.658–661 (2024 年 3 月).

(4) 東直輝, 大西俊輝, 木下峻一, 石井亮, 深山篤, 宮田章裕: マルチモーダル情報に基づく相槌種類予測の定性的評価, 情報処理学会研究報告コラボレーションとネットワークサービス (CN), Vol.2024-CN-121, No.29, pp.1–7 (2024 年 1 月).

(5) 東直輝, 大西俊輝, 木下駿一, 石井亮, 深山篤, 中村高雄, 宮田章裕: マルチモーダル情報に基づく多様な相槌の予測の検討, 情報処理学会シンポジウム論文集, マルチメディア、分散、協調とモバイル (DICOMO 2023), Vol.2023, pp.352–358 (2023 年 7 月).

(6) 木下峻一, 大西俊輝, 東直輝, 石井亮, 深山篤, 中村高雄, 大澤正彦, 宮田章裕: マルチモーダル情報に基づく聞き手の理解度推定の基礎検討, 情報処理学会研究報告グループウェアとネットワークサービス (GN), Vol.2023-GN-119, No.7, pp.1–6 (2023 年 3 月).

(7) 東直輝, 大西俊輝, 木下峻一, 石井亮, 深山篤, 中村高雄, 宮田章裕: マルチモーダル情報に基づく多様な相槌の生成の基礎検討, 情報処理学会研究報告グループウェアとネットワークサービス (GN), Vol.2023-GN-119, No.8, pp.1–6 (2023 年 3 月).

(8) 大串旭, 大西俊輝, 田原陽平, 石井亮, 深山篤, 中村高雄, 宮田章裕: 遠隔対話における上手い褒め方のモデリングの検討, 情報処理学会インタラクション 2023 論文集, pp.749–754 (2023 年 3 月).

(9) 大西俊輝, 木下峻一, 東直輝, 石井亮, 深山篤, 中村高雄, 宮田章裕: マルチモーダル情報に基づく聞き手のバックチャネルの種類推定の基礎検討, 情報処理学会グループウェアとネットワークサービスワークショップ 2022 論文集, Vol.2022, pp.64–66 (2022 年 11 月).

(10) 田原陽平, 大西俊輝, 大串旭, 石井亮, 深山篤, 中村高雄, 宮田章裕: マルチモーダル情報に基づく褒める行為の判定の基礎検討, 情報処理学会シンポジウム論文集, マルチメディア、分散、協調とモバイル (DICOMO 2022), Vol.2022, pp.576–581 (2022 年 7 月).

(11) 大串旭, 大西俊輝, 山内愛里沙, 石井亮, 杵渕哲也, 青野裕司, 宮田章裕: 言葉づかいに着目した褒め方の上手さの推定モデルの基礎検討, 情報処理学会シンポジウム論文

集, マルチメディア、分散、協調とモバイル (DICOMO 2021), Vol.2021, pp.791–797 (2021 年 7 月).

(12) 大串旭, 大西俊輝, 山内愛里沙, 石井亮, 杵渕哲也, 青野裕司, 宮田章裕: 上手く褒めるために効果的な言葉づかいの調査, 情報処理学会インタラクション 2021 論文集, pp.714–718 (2021 年 3 月).

(13) 山内愛里沙, 大西俊輝, 武藤佑太, 石井亮, 青野裕司, 宮田章裕: 音声および視線・表情・頭部運動に基づく上手い褒め方の評価システムの検討, 情報処理学会シンポジウム論文集, マルチメディア、分散、協調とモバイル (DICOMO 2020), Vol.2020, pp.98–106 (2020 年 7 月).

(14) 山内愛里沙, 大西俊輝, 呉健朗, 武藤佑太, 石井亮, 青野裕司, 宮田章裕: 表情・音声を用いた褒め方の上手さを評価するシステムの基礎検討, 情報処理学会インタラクション 2020 論文集, pp.247–252 (2020 年 3 月).

(15) 大西俊輝, 柴田万里那, 呉健朗, 石井亮, 富田準二, 宮田章裕: 対話における上手い褒め方のモデリングの基礎検討, 情報処理学会シンポジウム論文集, マルチメディア、分散、協調とモバイル (DICOMO 2019), Vol.2019, pp.656–662 (2019 年 7 月).

招待講演

(1) 大西俊輝: 非言語行動を用いた褒め方の上手さの分析, 第 22 回情報科学技術フォーラム (FIT2023), トップコンファレンスセッション (2023 年 9 月).

メディア掲載

(1) 日本大学新聞社, 日本大学新聞 (2 面): 情報処理学会「褒める行為」論文 ベストペーパー賞を受賞 (2023 年 3 月 17 日).

受賞

(1) マルチメディア、分散、協調とモバイル (DICOMO 2023) シンポジウム 優秀論文賞, マルチモーダル情報に基づく多様な相槌の予測の検討, 受賞者: 東直輝, 大西俊輝, 木下峻一, 石井亮, 深山篤, 中村高雄, 宮田章裕 (2023 年 9 月).

(2) 情報処理学会グループウェアとネットワークサービスワークショップ 2022 ベストペーパー賞, 対面・遠隔対話からの称賛行為検出の基礎検討, 受賞者: 田原陽平, 大西俊輝, 大串旭, 石井亮, 深山篤, 中村高雄, 宮田章裕 (2022 年 11 月).

(3) マルチメディア、分散、協調とモバイル (DICOMO 2022) シンポジウム 優秀論文賞, マルチモーダル情報に基づく褒める行為の判定の基礎検討, 受賞者: 田原陽平, 大西俊輝, 大串旭, 石井亮, 深山篤, 中村高雄, 宮田章裕 (2022 年 9 月).

(4) 情報処理学会グループウェアとネットワークサービスワークショップ 2021 ベストペーパー賞, 言語特徴に着目した褒め方の上手さの推定モデルの検討, 受賞者: 大串旭, 大西俊輝, 田原陽平, 石井亮, 深山篤, 中村高雄, 宮田章裕 (2021 年 11 月).

(5) マルチメディア、分散、協調とモバイル (DICOMO 2020) シンポジウム 優秀論文賞, 音声および視線・表情・頭部運動に基づく上手い褒め方の評価システムの検討, 受賞者: 山内愛里沙, 大西俊輝, 武藤佑太, 石井亮, 青野裕司, 宮田章裕 (2020 年 10 月).

(6) 情報処理学会グループウェアとネットワークサービスワークショップ 2019 ベストプレゼンテーション賞, 褒め方の上手さの推定における頭部・顔部の効果, 受賞者: 大西俊輝 (2019 年 11 月).

(7) マルチメディア、分散、協調とモバイル (DICOMO 2019) シンポジウム ヤングリサーチャ賞, 対話における上手い褒め方のモデリングの基礎検討, 受賞者: 大西俊輝 (2019 年 7 月).

本研究以外に関する業績

査読付き論文誌

(1) Kenro Go, Shunsuke Tokuda, Haruna Niiyama, Toshiki Onishi, Asahi Ogushi, and Akihiro Miyata: Evaluation of Prosodic Feature Suitable for Conversational Agents Replying with a Joke. Journal of Information Processing, Vol.32, pp.35–40 (Jan. 2024).

(2) Kenro Go, Toshiki Onishi, Asahi Ogushi, Shunsuke Tokuda, and Akihiro Miyata: Evaluation of Motivation for Conversational Agents Replying with Manzai-Style Jokes. Transactions of the Virtual Reality Society of Japan, Vol.28, No.1, pp.43–53 (Mar. 2023).

(3) 柴田万里那, 大西俊輝, 呉健朗, 宮田章裕: 柔らかい物体の動きによる共感表現方法の効果, 情報処理学会論文誌, Vol.62, No.1, pp.26–34 (2021 年 1 月).

(4) 立花巧樹, 呉健朗, 富永詩音, 大西俊輝, 鈴木颯馬, 諏訪博彦, 宮田章裕: 実世界オブジェクトを用いた生活空間内における事故予測支援手法, 情報処理学会論文誌, Vol.62, No.1, pp.35–43 (2021 年 1 月).

査読付き国際会議

(1) Haruna Niiyama, Toshiki Onishi, Kenro Go, Asahi Ogushi, and Akihiro Miyata: Does the Number of Agents with Vagueness Affect Empathy Expression? Proceedings of the 12th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW 2024), pp.210–213 (Sep. 2024).

(2) Kenro Go, Toshiki Onishi, Asahi Ogushi, and Akihiro Miyata: Conversational Agents Replying with a Manzai-style Joke. Proceedings of the 33rd Australian Conference on Human-computer-interaction (OzCHI 2021), pp.221–231 (Nov. 2021).

査読付き国内会議

(1) 新山はるな, 呉健朗, 大西俊輝, 大串旭, 宮田章裕: 2 体の柔らかい物体の動きによる共感表現方法の検証, 情報処理学会コラボレーションとネットワークサービスワークショップ 2023 論文集, Vol.2023, pp.24–32 (2023 年 11 月).

(2) 呉健朗, 大西俊輝, 大串旭, 宮田章裕: ノリツッコミを行う対話型エージェント, 情報処理学会インタラクション 2022 論文集, pp.39–47 (2022 年 2 月).

(3) 中原涼太, 長岡大二, 呉健朗, 大西俊輝, 柴田万里那, 宮田章裕: 複数対話型エージェントの役割分担によるユーモア生成システムの基礎検討, 情報処理学会グループウェアとネットワークサービスワークショップ 2018 論文集, Vol.2018, pp.1–8 (2018 年 11 月).

研究会・シンポジウム

(1) 渡貫健太, 新山はるな, 呉健朗, 大西俊輝, 大串旭, 宮田章裕: 非同期的な共感表現を行うエージェントの効果測定, 情報処理学会コラボレーションとネットワークサービスワークショップ 2024 論文集, Vol.2024, pp.63–70 (2024 年 11 月).

(2) 大西俊輝: ACII 2024 参加報告, 情報処理学会コラボレーションとネットワークサービスワークショップ 2024 論文集, Vol.2024, pp.173–176 (2024 年 11 月).

(3) 関一樹, 大西俊輝, 呉健朗, 木下峻一, 福田聡子, 奥岡耕平, 宮田章裕, 大澤正彦: 「準」自然言語エージェントを用いた寄付行動の促進, 日本認知科学会第 41 回大会, pp.519–522 (2024 年 10 月).

(4) 新山はるな, 呉健朗, 大西俊輝, 大串旭, 宮田章裕: 2 体の柔らかい物体の動きを用いた共感表現の認知に及ぼす共感スキルの影響調査, 情報処理学会シンポジウム論文集, マルチメディア、分散、協調とモバイル (DICOMO 2024), Vol.2024, pp.1211–1217 (2024 年 6 月).

(5) 丸山葉, 大西俊輝, 大串旭, 石井亮, 深山篤, 大澤正彦, 宮田章裕: デジタルツインを用いた自己効力感向上システムの実装, 情報処理学会シンポジウム論文集, マルチメディア、分散、協調とモバイル (DICOMO 2024), Vol.2024, pp.1579–1584 (2024 年 6 月).

(6) 渡貫健太, 呉健朗, 新山はるな, 大西俊輝, 大串旭, 宮田章裕: 語尾でボケて返す対話型エージェントの実装, 情報処理学会インタラクション 2024 論文集, pp.855–858 (2024 年 3 月).

(7) 渡辺好汰, 丸山葉, 大西俊輝, 大串旭, 呉健朗, 大澤正彦, 宮田章裕: 対話型エージェントが意見伝達を仲介する際に適した音声の高さに関する調査, 情報処理学会インタラクション 2024 論文集, pp.1429–1433 (2024 年 3 月).

(8) 丸山葉, 大西俊輝, 大串旭, 木下峻一, 石井亮, 深山篤, 大澤正彦, 宮田章裕: デジタルツインを用いた自己効力感向上システムの基礎検討, 情報処理学会コラボレーションとネットワークサービスワークショップ 2023 論文集, Vol.2023, pp.88–89 (2023 年 11 月).

(9) 大西俊輝: IVA 2023 参加報告, 情報処理学会コラボレーションとネットワークサービスワークショップ 2023 論文集, Vol.2023, pp.144–147 (2023 年 11 月).

(10) 呉健朗, 渡貫健太, 新山はるな, 大西俊輝, 大串旭, 宮田章裕: 語尾でボケて返す対話型エージェントの基礎検討, 情報処理学会コラボレーションとネットワークサービスワークショップ 2023 論文集, Vol.2023, pp.152–153 (2023 年 11 月).

(11) 新山はるな, 呉健朗, 大西俊輝, 大串旭, 宮田章裕: ユーモラスに話題提供を行うエージェントの検証, 情報処理学会シンポジウム論文集, マルチメディア、分散、協調とモバイル (DICOMO 2023), Vol.2023, pp.700–707 (2023 年 7 月).

(12) 大西俊輝, 新山はるな, 丸山葉, 大串旭, 呉健朗, 宮田章裕: 柔らかい物体の動きによる共感表現方法に及ぼす共感スキルの影響調査, 情報処理学会シンポジウム論文集, マルチメディア、分散、協調とモバイル (DICOMO 2023), Vol.2023, pp.708–715 (2023 年 7 月).

(13) 新山はるな, 得田舜介, 大串旭, 大西俊輝, 呉健朗, 大澤正彦, 宮田章裕: ユーモラスに話題提供を行うエージェントの基礎検討, 情報処理学会インタラクション 2023 論文集, pp.447–450 (2023 年 3 月).

(14) 関一樹, 内村方哉, 大西俊輝, 呉健朗, 宮田章裕, 大澤正彦: 寄付行動を促進する非自然言語エージェントの基礎検討, 情報処理学会グループウェアとネットワークサービスワークショップ 2022 論文集, Vol.2022, pp.141–146 (2022 年 11 月).

(15) 丸山葉, 大西俊輝, 大串旭, 呉健朗, 大澤正彦, 宮田章裕: 意見伝達を仲介する対話型エージェントに対する利用意欲の調査, 情報処理学会シンポジウム論文集, マルチメディア、分散、協調とモバイル (DICOMO 2022), Vol.2022, pp.369–373 (2022 年 7 月).

(16) 得田舜介, 呉健朗, 大西俊輝, 大串旭, 宮田章裕: 対話型エージェントのユーモア表現に適した韻律的特徴の調査, 情報処理学会シンポジウム論文集, マルチメディア、分散、協調とモバイル (DICOMO 2022), Vol.2022, pp.570–575 (2022 年 7 月).

(17) 丸山葉, 大西俊輝, 大串旭, 呉健朗, 大澤正彦, 宮田章裕: 意見伝達を仲介する対話型エージェントの基礎検討, HAI シンポジウム 2022 予稿集, G-17 (2022 年 3 月).

(18) 得田舜介, 呉健朗, 田中柊羽, 大西俊輝, 大串旭, 宮田章裕: 英語でボケる発言を行うエージェントの基礎検討, 情報処理学会インタラクション 2022 論文集, pp.291–294 (2022 年 2 月).

(19) 田中柊羽, 呉健朗, 大西俊輝, 大串旭, 武藤佑太, 宮田章裕: たとえツッコミを行う対話型エージェントの基礎検討, 情報処理学会インタラクション 2021 論文集, pp.639–641 (2021 年 3 月).

(20) 立花巧樹, 大西俊輝, 鈴木颯馬, 富永詩音, 呉健朗, 宮田章裕: 生活空間における危険予測支援システムの基礎検討, 情報処理学会グループウェアとネットワークサービスワークショップ 2019 論文集, Vol.2019, pp.99–102 (2019 年 11 月).

(21) 立花巧樹, 富永詩音, 大西俊輝, 呉健朗, 宮田章裕: ベビーカー利用時における周囲への動作予告システムの実装, 情報処理学会シンポジウム論文集, マルチメディア、分散、協調とモバイル (DICOMO 2019), Vol.2019, pp.1273–1279 (2019 年 7 月).

(22) 柴田万里那, 大西俊輝, 呉健朗, 宮田章裕: 柔らかい物体の動きによる共感表現方法の基礎検証, 情報処理学会研究報告, ユビキタスコンピューティングシステム (UBI), Vol.2019-UBI-62, No.11, pp.1–6 (2019 年 6 月).

(23) 立花巧樹, 富永詩音, 大西俊輝, 呉健朗, 宮田章裕: ベビーカー利用時における周囲への動作予告手法の基礎検討, 情報処理学会インタラクション 2019 論文集, pp.179–181 (2019 年 3 月).

(24) 柴田万里那, 大西俊輝, 呉健朗, 長岡大二, 中原涼太, 宮田章裕: 柔らかい物体の動きによる共感表現方法の基礎検討, 情報処理学会インタラクション 2019 論文集, pp.572–575 (2019 年 3 月).

受賞

(1) 情報処理学会コラボレーションとネットワークサービスワークショップ 2023 ベストペーパー賞, 2 体の柔らかい物体の動きによる共感表現方法の検証, 受賞者: 新山はるな, 呉健朗, 大西俊輝, 大串旭, 宮田章裕 (2022 年 11 月).

(2) 情報処理学会グループウェアとネットワークサービスワークショップ 2018 ベストペーパー賞, 複数対話型エージェントの役割分担によるユーモア生成システムの基礎検討, 受賞者: 中原涼太, 長岡大二, 呉健朗, 大西俊輝, 柴田万里那, 宮田章裕 (2018 年 11 月).